

CLP-2 INTEGRAL CALCULUS



*Joel Feldman, Andrew Rechnitzer & Elyse
Yeager*
University of British Columbia

University of British Columbia
CLP-2 Integral Calculus

Joel Feldman, Andrew Rechnitzer and Elyse
Yeager

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (<https://LibreTexts.org>) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by [NICE CXone Expert](#) and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptations contact info@LibreTexts.org or visit our main website at <https://LibreTexts.org>.

This text was compiled on 06/09/2025

TABLE OF CONTENTS

Colophon

Licensing

Preface

Dedication

Acknowledgements

Using the exercises in this book

Feedback about the text

1: Integration

- 1.1: Definition of the Integral
- 1.2: Basic properties of the definite integral
- 1.3: The Fundamental Theorem of Calculus
- 1.4: Substitution
- 1.5: Area between curves
- 1.6: Volumes
- 1.7: Integration by parts
- 1.8: Trigonometric Integrals
- 1.9: Trigonometric Substitution
- 1.10: Partial Fractions
- 1.11: Numerical Integration
- 1.12: Improper Integrals
- 1.13: More Integration Examples

2: Applications of Integration

- 2.1: Work
- 2.2: Averages
- 2.3: Center of Mass and Torque
- 2.4: Separable Differential Equations

3: Sequence and series

- 3.1: Sequences
- 3.2: Series
- 3.3: Convergence Tests
- 3.4: Absolute and Conditional Convergence
- 3.5: Power Series
- 3.6: Taylor Series
- 3.7: Optional — Rational and irrational numbers

Appendices

- A: High School Material
 - A.1: Similar Triangles

- [A.2: Pythagoras](#)
- [A.3: Trigonometry — Definitions](#)
- [A.4: Radians, Arcs and Sectors](#)
- [A.5: Trigonometry — Graphs](#)
- [A.6: Trigonometry — Special Triangles](#)
- [A.7: Trigonometry — Simple Identities](#)
- [A.8: Trigonometry — Add and Subtract Angles](#)
- [A.9: Inverse Trigonometric Functions](#)
- [A.10: Areas](#)
- [A.11: Volumes](#)
- [A.12: Powers](#)
- [A.13: Logarithms](#)
- [A.14: Highschool Material You Should be Able to Derive](#)
- [A.15: Cartesian Coordinates](#)
- [A.16: Roots of Polynomials](#)
- [B: Complex Numbers and Exponentials](#)
 - [B.1: Definition and Basic Operations](#)
 - [B.2: The Complex Exponential](#)
- [C: More About Numerical Integration](#)
 - [C.1: Richardson Extrapolation](#)
 - [C.2: Romberg Integration](#)
 - [C.3: Adaptive Quadrature](#)
- [D: Numerical Solution of ODE's](#)
 - [D.1: Simple ODE Solvers — Derivation](#)
 - [D.2: Simple ODE Solvers — Error Behaviour](#)
 - [D.3: Variable Step Size Methods](#)
- [E: Hints for Exercises](#)
- [F: Answers to Exercises](#)
- [G: Solutions to Exercises](#)

[Index](#)

[Glossary](#)

[Detailed Licensing](#)

Colophon

Cover Design: Nick Loewen — licensed under the [CC-BY-NC-SA 4.0 License](#).

Edition: CLP-2 Integral Calculus: May 2021

Website: CLP -2

©2016 – 2021 Joel Feldman, Andrew Rechnitzer, Elyse Yeager

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. You can view a copy of the license [here](#).



Source files A link to the source files for this document can be found at the [CLP textbook website](#). The sources are licensed under the [CC-BY-NC-SA 4.0 License](#).

Licensing

A detailed breakdown of this resource's licensing can be found in [Back Matter/Detailed Licensing](#).

Preface

This text is a merger of the CLP Integral Calculus textbook and problembook. It is, at the time that we write this, still a work in progress; some bits and pieces around the edges still need polish. Consequently we recommend to the student that they still consult text webpage for links to the errata — especially if they think there might be a typo or error. We also request that you send us an email at clp@ugrad.math.ubc.ca

Additionally, if you are not a student at UBC and using these texts please send us an email (again using the feedback button) — we'd love to hear from you.

Joel Feldman, Andrew Rechnitzer and Elyse Yeager

Dedication

To our students.

And to the many generations of scholars who have freely shared all this knowledge with us.

Acknowledgements

Elyse would like to thank her husband Seçkin Demirbaş for his endless patience, tireless support, and insightful feedback.

Andrew would like to thank Zoe for her patience and support while he hacked and also Locke for his insightful subediting. He should also thank the various coauthors for not being too impatient while other manuscripts were neglected.

Joel would like to thank a range of people for their support and encouragement over many years.

Thanks to

- Rob Beezer and David Farmer for their help converting this book from LATEX to this online PreTeXt format.
- Nick Loewen for designing the cover art, help with figures, colours, spelling and many discussions.
- The many people who have collaborated over the last couple of decades making exams and tests for first year calculus courses at UBC Mathematics. A great many of the exercises in the text come from questions in those tests and exams.

Finally, we'd like to thank those students who reported typos and errors they found in the text. Many of these students did so through our “bug bounty” program which was supported by the [Department of Mathematics](#), [Skylight](#) and the [Loafe Cafe](#) all at UBC.

Using the exercises in this book

Each problem in this book is split into four parts: Question, Hint, Answer, and Solution. As you are working problems, resist the temptation to prematurely peek at the hint or to click through to the answers and solutions in the appendix! It's important to allow yourself to struggle for a time with the material. Even professional mathematicians don't always know right away how to solve a problem. The art is in gathering your thoughts and figuring out a strategy to use what you know to find out what you don't.

If you find yourself at a real impasse, go ahead and look at the linked hint. Think about it for a while, and don't be afraid to read back in the notes to look for a key idea that will help you proceed. If you still can't solve the problem, well, we included the Solutions section for a reason! As you're reading the solutions, try hard to understand why we took the steps we did, instead of memorizing step-by-step how to solve that one particular problem.

If you struggled with a question quite a lot, it's probably a good idea to return to it in a few days. That might have been enough time for you to internalize the necessary ideas, and you might find it easily conquerable. Pat yourself on the back — sometimes math makes you feel good! If you're still having troubles, read over the solution again, with an emphasis on understanding why each step makes sense.

One of the reasons so many students are required to study calculus is the hope that it will improve their problem-solving skills. In this class, you will learn lots of concepts, and be asked to apply them in a variety of situations. Often, this will involve answering one really big problem by breaking it up into manageable chunks, solving those chunks, then putting the pieces back together. When you see a particularly long question, remain calm and look for a way to break it into pieces you can handle.

- **Working with Friends**

Study buddies are fantastic! If you don't already have friends in your class, you can ask your neighbours in lecture to form a group. Often, a question that you might bang your head against for an hour can be easily cleared up by a friend who sees what you've missed. Regular study times make sure you don't procrastinate too much, and friends help you maintain a positive attitude when you might otherwise succumb to frustration. Struggle in mathematics is desirable, but suffering is not.

When working in a group, make sure you try out problems on your own before coming together to discuss with others. Learning is a process, and getting answers to questions that you haven't considered on your own can rob you of the practice you need to master skills and concepts, and the tenacity you need to develop to become a competent problem-solver.

- **Types of Questions**

Questions outlined by a blue box make up the *representative question set*. This set of questions is intended to cover the most essential ideas in each section. These questions are usually highly typical of what you'd see on an exam, although some of them are atypical but carry an important moral. If you find yourself unconfident with the idea behind one of these, it's probably a good idea to practice similar questions.

This representative question set is our suggestion for a minimal selection of questions to work on. You are highly encouraged to work on more.

In addition to original problems, this book contains problems pulled from quizzes and exams given at UBC for Math 101 (first-semester calculus) and Math 121 (honours first-semester calculus). These problems are marked by “(*)”. The authors would like to acknowledge the contributions of the many people who collaborated to produce these exams over the years.

Finally, the questions are organized into three types: **Stage 1**, **Stage 2**, and **Stage 3**.

- Exercises — Stage 1

The first category is meant to test and improve your understanding of basic underlying concepts. These often do not involve much calculation. They range in difficulty from very basic reviews of definitions to questions that require you to be thoughtful about the concepts covered in the section.

- Exercises — Stage 2

Questions in this category are for practicing skills. It's not enough to understand the philosophical grounding of an idea: you have to be able to apply it in appropriate situations. This takes practice!

- Exercises — Stage 3

The last questions in each section go a little farther than “**Stage 2**”. Often they will combine more than one idea, incorporate review material, or ask you to apply your understanding of a concept to a new situation.

In exams, as in life, you will encounter questions of varying difficulty. A good skill to practice is recognizing the level of difficulty a problem poses. Exams will have some easy question, some standard questions, and some harder questions.

Feedback about the text

The CLP-2 Integral Calculus text is still undergoing testing and changes. Because of this we request that if you find a problem or error in the text then:

1. Please check the errata list that can be found at the text webpage.
2. Is the problem in the online version or the PDF version or both?
3. Note the URL of the online version and the page number in the PDF
4. Send an email to `clp @ugrad.math.ubc.ca` . Please be sure to include
 - a description of the error
 - the URL of the page, if found in the online edition
 - and if the problem also exists in the PDF, then the page number in the PDF and the compile date on the front page of PDF.

CHAPTER OVERVIEW

1: Integration

Calculus is built on two operations — differentiation and integration.

- Differentiation — as we saw last term, differentiation allows us to compute and study the instantaneous rate of change of quantities. At its most basic it allows us to compute tangent lines and velocities, but it also led us to quite sophisticated applications including approximation of functions through Taylor polynomials and optimisation of quantities by studying critical and singular points.
- Integration — at its most basic, allows us to analyse the area under a curve. Of course, its application and importance extend far beyond areas and it plays a central role in solving differential equations.



It is not immediately obvious that these two topics are related to each other. However, as we shall see, they are indeed intimately linked.

- [1.1: Definition of the Integral](#)
- [1.2: Basic properties of the definite integral](#)
- [1.3: The Fundamental Theorem of Calculus](#)
- [1.4: Substitution](#)
- [1.5: Area between curves](#)
- [1.6: Volumes](#)
- [1.7: Integration by parts](#)
- [1.8: Trigonometric Integrals](#)
- [1.9: Trigonometric Substitution](#)
- [1.10: Partial Fractions](#)
- [1.11: Numerical Integration](#)
- [1.12: Improper Integrals](#)
- [1.13: More Integration Examples](#)

This page titled [1: Integration](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.1: Definition of the Integral

Arguably the easiest way to introduce integration is by considering the area between the graph of a given function and the x -axis, between two specific vertical lines — such as is shown in the figure above. We'll follow this route by starting with a motivating example.

A Motivating Example

Let us find the area under the curve $y = e^x$ (and above the x -axis) for $0 \leq x \leq 1$. That is, the area of $\{(x, y) \mid 0 \leq y \leq e^x, 0 \leq x \leq 1\}$.



This area is equal to the “definite integral”

$$\text{Area} = \int_0^1 e^x dx$$

Do not worry about this notation or terminology just yet. We discuss it at length below. In different applications this quantity will have different interpretations — not just area. For example, if x is time and e^x is your velocity at time x , then we'll see later (in Example 1.1.18) that the specified area is the net distance travelled between time 0 and time 1. After we finish with the example, we'll mimic it to give a general definition of the integral $\int_a^b f(x) dx$.

✓ Example 1.1.1 Computing an area with vertical strips

We wish to compute the area of $\{(x, y) \mid 0 \leq y \leq e^x, 0 \leq x \leq 1\}$. We know, from our experience with e^x in differential calculus, that the curve $y = e^x$ is not easily written in terms of other simpler functions, so it is very unlikely that we would be able to write the area as a combination of simpler geometric objects such as triangles, rectangles or circles.

So rather than trying to write down the area exactly, our strategy is to approximate the area and then make our approximation more and more precise ¹. We choose ² to approximate the area as a union of a large number of tall thin (vertical) rectangles. As we take more and more rectangles we get better and better approximations. Taking the limit as the number of rectangles goes to infinity gives the exact area ³.

As a warm up exercise, we'll now just use four rectangles. In Example 1.1.2, below, we'll consider an arbitrary number of rectangles and then take the limit as the number of rectangles goes to infinity. So

- subdivide the interval $0 \leq x \leq 1$ into 4 equal subintervals each of width $\frac{1}{4}$, and
- subdivide the area of interest into four corresponding vertical strips, as in the figure below.

The area we want is exactly the sum of the areas of all four strips.



Each of these strips is almost, but not quite, a rectangle. While the bottom and sides are fine (the sides are at right-angles to the base), the top of the strip is not horizontal. This is where we must start to approximate. We can replace each strip by a rectangle by just levelling off the top. But now we have to make a choice — at what height do we level off the top?

Consider, for example, the leftmost strip. On this strip, x runs from 0 to $\frac{1}{4}$. As x runs from 0 to $\frac{1}{4}$, the height y runs from e^0 to $e^{\frac{1}{4}}$. It would be reasonable to choose the height of the approximating rectangle to be somewhere between e^0 and $e^{\frac{1}{4}}$. Which



height should we choose? Well, actually it doesn't matter. When we eventually take the limit of infinitely many approximating rectangles all of those different choices give exactly the same final answer. We'll say more about this later.

In this example we'll do two sample computations.

- For the first computation we approximate each slice by a rectangle whose height is the height of the *left* hand side of the slice.
 - On the first slice, x runs from 0 to $\frac{1}{4}$, and the height y runs from e^0 , on the left hand side, to $e^{\frac{1}{4}}$, on the right hand side.
 - So we approximate the first slice by the rectangle of height e^0 and width $\frac{1}{4}$, and hence of area $\frac{1}{4} e^0 = \frac{1}{4}$.

- On the second slice, x runs from $\frac{1}{4}$ to $\frac{1}{2}$, and the height y runs from $e^{\frac{1}{4}}$ and $e^{\frac{1}{2}}$.
- So we approximate the second slice by the rectangle of height $e^{\frac{1}{4}}$ and width $\frac{1}{4}$, and hence of area $\frac{1}{4} e^{\frac{1}{4}}$.
- And so on.
- All together, we approximate the area of interest by the sum of the areas of the four approximating rectangles, which is

$$\left[1 + e^{\frac{1}{4}} + e^{\frac{1}{2}} + e^{\frac{3}{4}}\right] \frac{1}{4} = 1.5124$$

- This particular approximation is called the “left Riemann sum approximation to $\int_0^1 e^x dx$ with 4 subintervals”. We’ll explain this terminology later.
- This particular approximation represents the shaded area in the figure on the left below. Note that, because e^x increases as x increases, this approximation is definitely smaller than the true area.



- For the second computation we approximate each slice by a rectangle whose height is the height of the *right* hand side of the slice.
 - On the first slice, x runs from 0 to $\frac{1}{4}$, and the height y runs from e^0 , on the left hand side, to $e^{\frac{1}{4}}$, on the right hand side.
 - So we approximate the first slice by the rectangle of height $e^{\frac{1}{4}}$ and width $\frac{1}{4}$, and hence of area $\frac{1}{4} e^{\frac{1}{4}}$.
 - On the second slice, x runs from $\frac{1}{4}$ to $\frac{1}{2}$, and the height y runs from $e^{\frac{1}{4}}$ and $e^{\frac{1}{2}}$.
 - So we approximate the second slice by the rectangle of height $e^{\frac{1}{2}}$ and width $\frac{1}{4}$, and hence of area $\frac{1}{4} e^{\frac{1}{2}}$.
 - And so on.
 - All together, we approximate the area of interest by the sum of the areas of the four approximating rectangles, which is

$$\left[e^{\frac{1}{4}} + e^{\frac{1}{2}} + e^{\frac{3}{4}} + e^1\right] \frac{1}{4} = 1.9420$$

- This particular approximation is called the “right Riemann sum approximation to $\int_0^1 e^x dx$ with 4 subintervals”.
- This particular approximation represents the shaded area in the figure on the right above. Note that, because e^x increases as x increases, this approximation is definitely larger than the true area.

Now for the full computation that gives the exact area.

✓ Example 1.1.2 Computing an area exactly

Recall that we wish to compute the area of

$$\{(x, y) \mid 0 \leq y \leq e^x, 0 \leq x \leq 1\}$$

and that our strategy is to approximate this area by the area of a union of a large number of very thin rectangles, and then take the limit as the number of rectangles goes to infinity. In Example 1.1.1, we used just four rectangles. Now we’ll consider a general number of rectangles, that we’ll call n . Then we’ll take the limit $n \rightarrow \infty$. So

- pick a natural number n and
- subdivide the interval $0 \leq x \leq 1$ into n equal subintervals each of width $\frac{1}{n}$, and
- subdivide the area of interest into corresponding thin strips, as in the figure below.

The area we want is exactly the sum of the areas of all of the thin strips.



Each of these strips is almost, but not quite, a rectangle. As in Example 1.1.1, the only problem is that the top is not horizontal. So we approximate each strip by a rectangle, just by levelling off the top. Again, we have to make a choice — at what height do we level off the top?

Consider, for example, the leftmost strip. On this strip, x runs from 0 to $\frac{1}{n}$. As x runs from 0 to $\frac{1}{n}$, the height y runs from e^0 to $e^{\frac{1}{n}}$. It would be reasonable to choose the height of the approximating rectangle to be somewhere between e^0 and $e^{\frac{1}{n}}$. Which height should we choose?

Well, as we said in Example 1.1.1, it doesn't matter. We shall shortly take the limit $n \rightarrow \infty$ and, in that limit, all of those different choices give exactly the same final answer. We won't justify that statement in this example, but there will be an (optional) section shortly that provides the justification. For this example we just, arbitrarily, choose the height of each rectangle to be the height of the graph $y = e^x$ at the smallest value of x in the corresponding strip⁴. The figure on the left below shows the approximating rectangles when $n = 4$ and the figure on the right shows the approximating rectangles when $n = 8$.



Now we compute the approximating area when there are n strips.

- We approximate the leftmost strip by a rectangle of height e^0 . All of the rectangles have width $\frac{1}{n}$. So the leftmost rectangle has area $\frac{1}{n}e^0$.
- On strip number 2, x runs from $\frac{1}{n}$ to $\frac{2}{n}$. So the smallest value of x on strip number 2 is $\frac{1}{n}$, and we approximate strip number 2 by a rectangle of height $e^{\frac{1}{n}}$ and hence of area $\frac{1}{n}e^{\frac{1}{n}}$.
- And so on.
- On the last strip, x runs from $\frac{n-1}{n}$ to $\frac{n}{n} = 1$. So the smallest value of x on the last strip is $\frac{n-1}{n}$, and we approximate the last strip by a rectangle of height $e^{\frac{(n-1)}{n}}$ and hence of area $\frac{1}{n}e^{\frac{(n-1)}{n}}$.

The total area of all of the approximating rectangles is

$$\begin{aligned} \text{Total approximating area} &= \frac{1}{n}e^0 + \frac{1}{n}e^{\frac{1}{n}} + \frac{1}{n}e^{\frac{2}{n}} + \frac{1}{n}e^{\frac{3}{n}} + \cdots + \frac{1}{n}e^{\frac{(n-1)}{n}} \\ &= \frac{1}{n} \left(1 + e^{\frac{1}{n}} + e^{\frac{2}{n}} + e^{\frac{3}{n}} + \cdots + e^{\frac{(n-1)}{n}} \right) \end{aligned}$$

Now the sum in the brackets might look a little intimidating because of all the exponentials, but it actually has a pretty simple structure that can be easily seen if we rename $e^{\frac{1}{n}} = r$. Then

- the first term is $1 = r^0$ and
- the second term is $e^{\frac{1}{n}} = r^1$ and
- the third term is $e^{\frac{2}{n}} = r^2$ and
- the fourth term is $e^{\frac{3}{n}} = r^3$ and
- and so on and
- the last term is $e^{\frac{(n-1)}{n}} = r^{n-1}$.

So

$$\text{Total approximating area} = \frac{1}{n} (1 + r + r^2 + \cdots + r^{n-1})$$

The sum in brackets is known as a geometric sum and satisfies a nice simple formula:

Equation 1.1.3 Geometric sum

$$1 + r + r^2 + \cdots + r^{n-1} = \frac{r^n - 1}{r - 1} \quad \text{provided } r \neq 1$$

The derivation of the above formula is not too difficult. So let's derive it in a little aside.

Geometric sum

Denote the sum as

$$S = 1 + r + r^2 + \cdots + r^{n-1}$$

Notice that if we multiply the whole sum by r we get back almost the same thing:

$$\begin{aligned}
 rS &= r(1 + r + \dots + r^{n-1}) \\
 &= r + r^2 + r^3 + \dots + r^n
 \end{aligned}$$

This right hand side differs from the original sum S only in that

- the right hand side, which starts with “ $r +$ ”, is missing the “ $1 +$ ” that S starts with, and
- the right hand side has an extra “ $+r^n$ ” at the end that does not appear in S .

That is

$$rS = S - 1 + r^n$$

Moving this around a little gives

$$\begin{aligned}
 (r-1)S &= (r^n - 1) \\
 S &= \frac{r^n - 1}{r - 1}
 \end{aligned}$$

as required. Notice that the last step in the manipulations only works providing $r \neq 1$ (otherwise we are dividing by zero).

Now we can go back to our area approximation armed with the above result about geometric sums.

$$\begin{aligned}
 \text{Total approximating area} &= \frac{1}{n}(1 + r + r^2 + \dots + r^{n-1}) \\
 &= \frac{1}{n} \frac{r^n - 1}{r - 1} && \text{remember that } r = e^{1/n} \\
 &= \frac{1}{n} \frac{e^{n/n} - 1}{e^{1/n} - 1} \\
 &= \frac{1}{n} \frac{e - 1}{e^{1/n} - 1}
 \end{aligned}$$

To get the exact area⁵ all we need to do is make the approximation better and better by taking the limit $n \rightarrow \infty$. The limit will look more familiar if we rename $\frac{1}{n}$ to X . As n tends to infinity, X tends to 0, so

$$\begin{aligned}
 \text{Area} &= \lim_{n \rightarrow \infty} \frac{1}{n} \frac{e - 1}{e^{1/n} - 1} \\
 &= (e - 1) \lim_{n \rightarrow \infty} \frac{1/n}{e^{1/n} - 1} \\
 &= (e - 1) \lim_{X \rightarrow 0} \frac{X}{e^X - 1} \quad \left(\text{with } X = \frac{1}{n}\right)
 \end{aligned}$$

Examining this limit we see that both numerator and denominator tend to zero as $X \rightarrow 0$, and so we cannot evaluate this limit by computing the limits of the numerator and denominator separately and then dividing the results. Despite this, the limit is not too hard to evaluate; here we give two ways:

- Perhaps the easiest way to compute the limit is by using l'Hôpital's rule⁶. Since both numerator and denominator go to zero, this is a $\frac{0}{0}$ indeterminate form. Thus

$$\lim_{X \rightarrow 0} \frac{X}{e^X - 1} = \lim_{X \rightarrow 0} \frac{\frac{d}{dX} X}{\frac{d}{dX} (e^X - 1)} = \lim_{X \rightarrow 0} \frac{1}{e^X} = 1$$

- Another way⁷ to evaluate the same limit is to observe that it can be massaged into the form of the limit definition of the derivative. First notice that

$$\lim_{X \rightarrow 0} \frac{X}{e^X - 1} = \left[\lim_{X \rightarrow 0} \frac{e^X - 1}{X} \right]^{-1}$$

provided this second limit exists and is nonzero⁸. This second limit should look a little familiar:

$$\lim_{X \rightarrow 0} \frac{e^X - 1}{X} = \lim_{X \rightarrow 0} \frac{e^X - e^0}{X - 0}$$

which is just the definition of the derivative of e^x at $x = 0$. Hence we have

$$\begin{aligned} \lim_{X \rightarrow 0} \frac{X}{e^X - 1} &= \left[\lim_{X \rightarrow 0} \frac{e^X - e^0}{X - 0} \right]^{-1} \\ &= \left[\frac{d}{dX} e^X \Big|_{X=0} \right]^{-1} \\ &= [e^X \Big|_{X=0}]^{-1} \\ &= 1 \end{aligned}$$

So, after this short aside into limits, we may now conclude that

$$\begin{aligned} \text{Area} &= (e - 1) \lim_{X \rightarrow 0} \frac{X}{e^X - 1} \\ &= e - 1 \end{aligned}$$

Optional — A more rigorous area computation

In Example 1.1.1 above we considered the area of the region $\{ (x, y) \mid 0 \leq y \leq e^x, 0 \leq x \leq 1 \}$. We approximated that area by the area of a union of n thin rectangles. We then claimed that upon taking the number of rectangles to infinity, the approximation of the area became the exact area. However we did not justify the claim. The purpose of this optional section is to make that calculation rigorous.

The broad set-up is the same. We divide the region up into n vertical strips, each of width $\frac{1}{n}$ and we then approximate those strips by rectangles. However rather than an uncontrolled approximation, we construct two sets of rectangles — one set always smaller than the original area and one always larger. This then gives us lower and upper bounds on the area of the region. Finally we make use of the squeeze theorem⁹ to establish the result.

- To find our upper and lower bounds we make use of the fact that e^x is an increasing function. We know this because the derivative $\frac{d}{dx} e^x = e^x$ is always positive. Consequently, the smallest and largest values of e^x on the interval $a \leq x \leq b$ are e^a and e^b , respectively.
- In particular, for $0 \leq x \leq \frac{1}{n}$, e^x takes values only between e^0 and $e^{\frac{1}{n}}$. As a result, the first strip

$$\{ (x, y) \mid 0 \leq x \leq \frac{1}{n}, 0 \leq y \leq e^x \}$$

- contains the rectangle of $0 \leq x \leq \frac{1}{n}$, $0 \leq y \leq e^0$ (the lighter rectangle in the figure on the left below) and
- is contained in the rectangle $0 \leq x \leq \frac{1}{n}$, $0 \leq y \leq e^{\frac{1}{n}}$ (the largest rectangle in the figure on the left below).

Hence

$$\frac{1}{n} e^0 \leq \text{Area} \{ (x, y) \mid 0 \leq x \leq \frac{1}{n}, 0 \leq y \leq e^x \} \leq \frac{1}{n} e^{\frac{1}{n}}$$



- Similarly, for the second, third, ..., last strips, as in the figure on the right above,

$$\frac{1}{n} e^{\frac{1}{n}} \leq \text{Area} \{ (x, y) \mid \frac{1}{n} \leq x \leq \frac{2}{n}, 0 \leq y \leq e^x \} \leq \frac{1}{n} e^{\frac{2}{n}}$$

$$\frac{1}{n} e^{\frac{2}{n}} \leq \text{Area} \{ (x, y) \mid \frac{2}{n} \leq x \leq \frac{3}{n}, 0 \leq y \leq e^x \} \leq \frac{1}{n} e^{\frac{3}{n}}$$

⋮

$$\frac{1}{n} e^{\frac{(n-1)}{n}} \leq \text{Area} \{ (x, y) \mid \frac{(n-1)}{n} \leq x \leq \frac{n}{n}, 0 \leq y \leq e^x \} \leq \frac{1}{n} e^{\frac{n}{n}}$$

- Adding these n inequalities together gives

$$\begin{aligned} & \frac{1}{n} \left(1 + e^{\frac{1}{n}} + \dots + e^{\frac{(n-1)}{n}} \right) \\ & \leq \text{Area} \{ (x, y) \mid 0 \leq x \leq 1, 0 \leq y \leq e^x \} \\ & \leq \frac{1}{n} \left(e^{\frac{1}{n}} + e^{\frac{2}{n}} + \dots + e^{\frac{n}{n}} \right) \end{aligned}$$

- We can then recycle equation 1.1.3 with $r = e^{\frac{1}{n}}$, so that $r^n = \left(e^{\frac{1}{n}} \right)^n = e$. Thus we have

$$\frac{1}{n} \frac{e-1}{e^{\frac{1}{n}}-1} \leq \text{Area} \{ (x, y) \mid 0 \leq x \leq 1, 0 \leq y \leq e^x \} \leq \frac{1}{n} e^{\frac{1}{n}} \frac{e-1}{e^{\frac{1}{n}}-1}$$

where we have used the fact that the upper bound is a simple multiple of the lower bound:

$$\left(e^{\frac{1}{n}} + e^{\frac{2}{n}} + \dots + e^{\frac{n}{n}} \right) = e^{\frac{1}{n}} \left(1 + e^{\frac{1}{n}} + \dots + e^{\frac{(n-1)}{n}} \right).$$

- We now apply the squeeze theorem to the above inequalities. In particular, the limits of the lower and upper bounds are $\lim_{n \rightarrow \infty} \frac{1}{n} \frac{e-1}{e^{\frac{1}{n}}-1}$ and $\lim_{n \rightarrow \infty} \frac{1}{n} e^{\frac{1}{n}} \frac{e-1}{e^{\frac{1}{n}}-1}$, respectively. As we did near the end of Example 1.1.2, we make these limits look more familiar by renaming $\frac{1}{n}$ to X . As n tends to infinity, X tends to 0, so the limits of the lower and upper bounds are

$$\lim_{n \rightarrow \infty} \frac{1}{n} \frac{e-1}{e^{\frac{1}{n}}-1} = (e-1) \lim_{X=\frac{1}{n} \rightarrow 0} \frac{X}{e^X-1} = e-1$$

(by l'Hôpital's rule) and

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} e^{\frac{1}{n}} \frac{e-1}{e^{\frac{1}{n}}-1} &= (e-1) \lim_{X=\frac{1}{n} \rightarrow 0} \frac{X e^X}{e^X-1} \\ &= (e-1) \lim_{X \rightarrow 0} e^X \cdot \lim_{X \rightarrow 0} \frac{X}{e^X-1} \\ &= (e-1) \cdot 1 \cdot 1 \end{aligned}$$

Thus, since the exact area is trapped between the lower and upper bounds, the squeeze theorem then implies that

$$\text{Exact area} = e - 1.$$

Summation notation

As you can see from the above example (and the more careful rigorous computation), our discussion of integration will involve a fair bit of work with sums of quantities. To this end, we make a quick aside into summation notation. While one can work through the material below without this notation, proper summation notation is well worth learning, so we advise the reader to persevere.

Writing out the summands explicitly can become quite impractical — for example, say we need the sum of the first 11 squares:

$$1 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 8^2 + 9^2 + 10^2 + 11^2$$

This becomes tedious. Where the pattern is clear, we will often skip the middle few terms and instead write

$$1 + 2^2 + \dots + 11^2.$$

A far more precise way to write this is using Σ (capital-sigma) notation. For example, we can write the above sum as

$$\sum_{k=1}^{11} k^2$$

This is read as

! The sum from k equals 1 to 11 of k^2 .

More generally

 Definition 1.1.4

Let $m \leq n$ be integers and let $f(x)$ be a function defined on the integers. Then we write

$$\sum_{k=m}^n f(k)$$

to mean the sum of $f(k)$ for k from m to n :

$$f(m) + f(m+1) + f(m+2) + \cdots + f(n-1) + f(n).$$

Similarly we write

$$\sum_{i=m}^n a_i$$

to mean

$$a_m + a_{m+1} + a_{m+2} + \cdots + a_{n-1} + a_n$$

for some set of coefficients $\{a_m, \dots, a_n\}$.

Consider the example

$$\sum_{k=3}^7 \frac{1}{k^2} = \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{7^2}$$

It is important to note that the right hand side of this expression evaluates to a number¹⁰; it does not contain “ k ”. The summation index k is just a “dummy” variable and it does not have to be called k . For example

$$\sum_{k=3}^7 \frac{1}{k^2} = \sum_{i=3}^7 \frac{1}{i^2} = \sum_{j=3}^7 \frac{1}{j^2} = \sum_{\ell=3}^7 \frac{1}{\ell^2}$$

Also the summation index has no meaning outside the sum. For example

$$k \sum_{k=3}^7 \frac{1}{k^2}$$

has no mathematical meaning; it is gibberish.

A sum can be represented using summation notation in many different ways. If you are unsure as to whether or not two summation notations represent the same sum, just write out the first few terms and the last couple of terms. For example,

$$\begin{aligned} \sum_{m=3}^{15} \frac{1}{m^2} &= \overbrace{\frac{1}{3^2}}^{m=3} + \overbrace{\frac{1}{4^2}}^{m=4} + \overbrace{\frac{1}{5^2}}^{m=5} + \cdots + \overbrace{\frac{1}{14^2}}^{m=14} + \overbrace{\frac{1}{15^2}}^{m=15} \\ \sum_{m=4}^{16} \frac{1}{(m-1)^2} &= \overbrace{\frac{1}{3^2}}^{m=4} + \overbrace{\frac{1}{4^2}}^{m=5} + \overbrace{\frac{1}{5^2}}^{m=6} + \cdots + \overbrace{\frac{1}{14^2}}^{m=15} + \overbrace{\frac{1}{15^2}}^{m=16} \end{aligned}$$

are equal.

Here is a theorem that gives a few rules for manipulating summation notation.

 Theorem 1.1.5 Arithmetic of Summation Notation

Let $n \geq m$ be integers. Then for all real numbers c and $a_i, b_i, m \leq i \leq n$.

- $\sum_{i=m}^n ca_i = c \left(\sum_{i=m}^n a_i \right)$

$$2. \sum_{i=m}^n (a_i + b_i) = \left(\sum_{i=m}^n a_i \right) + \left(\sum_{i=m}^n b_i \right)$$

$$3. \sum_{i=m}^n (a_i - b_i) = \left(\sum_{i=m}^n a_i \right) - \left(\sum_{i=m}^n b_i \right)$$

Proof

We can prove this theorem by just writing out both sides of each equation, and observing that they are equal, by the usual laws of arithmetic¹¹. For example, for the first equation, the left and right hand sides are

$$\sum_{i=m}^n ca_i = ca_m + ca_{m+1} + \cdots + ca_n$$

and $c \left(\sum_{i=m}^n a_i \right) = c(a_m + a_{m+1} + \cdots + a_n)$

They are equal by the usual distributive law. The “distributive law” is the fancy name for $c(a + b) = ca + cb$.

Not many sums can be computed exactly¹². Here are some that can. The first few are used a lot.

Theorem 1.1.6

1. $\sum_{i=0}^n ar^i = a \frac{1-r^{n+1}}{1-r}$, for all real numbers a and $r \neq 1$ and all integers $n \geq 0$.
2. $\sum_{i=1}^n 1 = n$, for all integers $n \geq 1$.
3. $\sum_{i=1}^n i = \frac{1}{2}n(n+1)$, for all integers $n \geq 1$.
4. $\sum_{i=1}^n i^2 = \frac{1}{6}n(n+1)(2n+1)$, for all integers $n \geq 1$.
5. $\sum_{i=1}^n i^3 = \left[\frac{1}{2}n(n+1) \right]^2$, for all integers $n \geq 1$.

Proof of Theorem 1.1.6 (Optional)

Proof

1. The first sum is

$$\sum_{i=0}^n ar^i = ar^0 + ar^1 + ar^2 + \cdots + ar^n$$

which is just the left hand side of equation 1.1.3, with n replaced by $n+1$ and then multiplied by a .

2. The second sum is just n copies of 1 added together, so of course the sum is n .
3. The third and fourth sums are discussed in the appendix of the CLP-1 text. In that discussion certain “tricks” are used to compute the sums with only simple arithmetic. Those tricks do not easily generalise to the fifth sum.
4. Instead of repeating that appendix, we’ll derive the third sum using a trick that generalises to the fourth and fifth sums (and also to higher powers). The trick uses the generating function¹³ $S(x)$:

Equation 1.1.7 Finite geometric sum

$$S(x) = 1 + x + x^2 + \cdots + x^n = \frac{x^{n+1} - 1}{x - 1}$$

Notice that this is just the geometric sum given by equation 1.1.3 with n replaced by $n+1$.

Now, consider the limit

$$\begin{aligned} \lim_{x \rightarrow 1} S(x) &= \lim_{x \rightarrow 1} (1 + x + x^2 + \dots + x^n) = n + 1 \quad \text{but also} \\ &= \lim_{x \rightarrow 1} \frac{x^{n+1} - 1}{x - 1} \quad \text{now use l'Hôpital's rule} \\ &= \lim_{x \rightarrow 1} \frac{(n+1)x^n}{1} = n + 1. \end{aligned}$$

This is not so hard (or useful). But now consider the derivative of $S(x)$:

$$\begin{aligned} S'(x) &= 1 + 2x + 3x^2 + \dots + nx^{n-1} \\ &= \frac{d}{dx} \left[\frac{x^{n+1} - 1}{x - 1} \right] \quad \text{use the quotient rule} \\ &= \frac{(x-1) \cdot (n+1)x^n - (x^{n+1} - 1) \cdot 1}{(x-1)^2} \quad \text{now clean it up} \\ &= \frac{nx^{n+1} - (n+1)x^n + 1}{(x-1)^2}. \end{aligned}$$

Hence if we take the limit of the above expression as $x \rightarrow 1$ we recover

$$\begin{aligned} \lim_{x \rightarrow 1} S'(x) &= 1 + 2 + 3 + \dots + n \\ &= \lim_{x \rightarrow 1} \frac{nx^{n+1} - (n+1)x^n + 1}{(x-1)^2} \quad \text{now use l'Hôpital's rule} \\ &= \lim_{x \rightarrow 1} \frac{n(n+1)x^n - n(n+1)x^{n-1}}{2(x-1)} \quad \text{l'Hôpital's rule again} \\ &= \lim_{x \rightarrow 1} \frac{n^2(n+1)x^{n-1} - n(n+1)(n-1)x^{n-2}}{2} \\ &= \frac{n^2(n+1) - n(n-1)(n+1)}{2} = \frac{n(n+1)}{2} \end{aligned}$$

as required. This computation can be done without l'Hôpital's rule, but the manipulations required are a fair bit messier.

5. The derivation of the fourth and fifth sums is similar to, but even more tedious than, that of the third sum. One takes two or three derivatives of the generating functional.

The Definition of the Definite Integral

In this section we give a definition of the definite integral $\int_a^b f(x) dx$ generalising the machinery we used in Example 1.1.1. But first some terminology and a couple of remarks to better motivate the definition.

Definition 1.1.8

The symbol $\int_a^b f(x) dx$ is read “the definite integral of the function $f(x)$ from a to b ”. The function $f(x)$ is called the integrand of $\int_a^b f(x) dx$ and a and b are called¹⁴ the limits of integration. The interval $a \leq x \leq b$ is called the interval of integration and is also called the domain of integration.

Before we explain more precisely what the definite integral actually is, a few remarks (actually — a few interpretations) are in order.

- If $f(x) \geq 0$ and $a \leq b$, one interpretation of the symbol $\int_a^b f(x) dx$ is “the area of the region $\{ (x, y) \mid a \leq x \leq b, 0 \leq y \leq f(x) \}$ ”.

In this way we can rewrite the area in Example 1.1.1 as the definite integral $\int_0^1 e^x dx$.

- This interpretation breaks down when either $a > b$ or $f(x)$ is not always positive, but it can be repaired by considering “signed areas”.
- If $a \leq b$, but $f(x)$ is not always positive, one interpretation of $\int_a^b f(x) dx$ is “the signed area between $y = f(x)$ and the x -axis for $a \leq x \leq b$ ”. For “signed area” (which is also called the “net area”), areas above the x -axis count as positive while areas below the x -axis count as negative. In the example below, we have the graph of the function

$$f(x) = \begin{cases} -1 & \text{if } 1 \leq x \leq 2 \\ 2 & \text{if } 2 < x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

The 2×2 shaded square above the x -axis has signed area $+2 \times 2 = +4$. The 1×1 shaded square below the x -axis has signed area $-1 \times 1 = -1$. So, for this $f(x)$,

$$\int_0^5 f(x) dx = +4 - 1 = 3$$

- We'll come back to the case $b < a$ later.

We're now ready to define $\int_a^b f(x) dx$. The definition is a little involved, but essentially mimics what we did in Example 1.1.1 (which is why we did the example before the definition). The main differences are that we replace the function e^x by a generic function $f(x)$ and we replace the interval from 0 to 1 by the generic interval¹⁵ from a to b .

- We start by selecting any natural number n and subdividing the interval from a to b into n equal subintervals. Each subinterval has width $\frac{b-a}{n}$.
- Just as was the case in Example 1.1.1 we will eventually take the limit as $n \rightarrow \infty$, which squeezes the width of each subinterval down to zero.
- For each integer $0 \leq i \leq n$, define $x_i = a + i \cdot \frac{b-a}{n}$. Note that this means that $x_0 = a$ and $x_n = b$. It is worth keeping in mind that these numbers x_i do depend on n even though our choice of notation hides this dependence.
- Subinterval number i is $x_{i-1} \leq x \leq x_i$. In particular, on the first subinterval, x runs from $x_0 = a$ to $x_1 = a + \frac{b-a}{n}$. On the second subinterval, x runs from x_1 to $x_2 = a + 2\frac{b-a}{n}$.
- On each subinterval we now pick $x_{i,n}^*$ between x_{i-1} and x_i . We then approximate $f(x)$ on the i^{th} subinterval by the constant function $y = f(x_{i,n}^*)$. We include n in the subscript to remind ourselves that these numbers depend on n .

Geometrically, we're approximating the region

$$\{ (x, y) \mid x \text{ is between } x_{i-1} \text{ and } x_i, \text{ and } y \text{ is between } 0 \text{ and } f(x) \}$$

by the rectangle

$$\{ (x, y) \mid x \text{ is between } x_{i-1} \text{ and } x_i, \text{ and } y \text{ is between } 0 \text{ and } f(x_{i,n}^*) \}$$

In Example 1.1.1 we chose $x_{i,n}^* = x_{i-1}$ and so we approximated the function e^x on each subinterval by the value it took at the leftmost point in that subinterval.

- So, when there are n subintervals our approximation to the signed area between the curve $y = f(x)$ and the x -axis, with x running from a to b , is

$$\sum_{i=1}^n f(x_{i,n}^*) \cdot \frac{b-a}{n}$$

We interpret this as the signed area since the summands $f(x_{i,n}^*) \cdot \frac{b-a}{n}$ need not be positive.

- Finally we define the definite integral by taking the limit of this sum as $n \rightarrow \infty$.

Oof! This is quite an involved process, but we can now write down the definition we need.

 Definition 1.1.9

Let a and b be two real numbers and let $f(x)$ be a function that is defined for all x between a and b . Then we define

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_{i,n}^*) \cdot \frac{b-a}{n}$$

when the limit exists and takes the same value for all choices of the $x_{i,n}^*$'s. In this case, we say that f is integrable on the interval from a to b .

Of course, it is not immediately obvious when this limit should exist. Thankfully it is easier for a function to be “integrable” than it is for it to be “differentiable”.

 Theorem 1.1.10

Let $f(x)$ be a function on the interval $[a, b]$. If

- $f(x)$ is continuous on $[a, b]$, or
- $f(x)$ has a finite number of jump discontinuities on $[a, b]$ (and is otherwise continuous)

then $f(x)$ is integrable on $[a, b]$.

We will not justify this theorem. But a slightly weaker statement is proved in (the optional) Section 1.1.7. Of course this does not tell us how to actually evaluate any definite integrals — but we will get to that in time.

Some comments:

- Note that, in Definition 1.1.9, we allow a and b to be any two real numbers. We do not require that $a < b$. That is, even when $a > b$, the symbol $\int_a^b f(x) dx$ is still defined by the formula of Definition 1.1.9. We'll get an interpretation for $\int_a^b f(x) dx$, when $a > b$, later.
- It is important to note that the definite integral $\int_a^b f(x) dx$ represents a number, not a function of x . The integration variable x is another “dummy” variable, just like the summation index i in $\sum_{i=m}^n a_i$ (see Section 1.1.3). The integration variable does not have to be called x . For example

$$\int_a^b f(x) dx = \int_a^b f(t) dt = \int_a^b f(u) du$$

Just as with summation variables, the integration variable x has no meaning outside of $f(x) dx$. For example

$$x \int_0^1 e^x dx \quad \text{and} \quad \int_0^x e^x dx$$

are both gibberish.

The sum inside definition 1.1.9 is named after Bernhard Riemann¹⁶ who made the first rigorous definition of the definite integral and so placed integral calculus on rigorous footings.

 Definition 1.1.11

The sum inside definition 1.1.9

$$\sum_{i=1}^n f(x_{i,n}^*) \frac{b-a}{n}$$

is called a Riemann sum. It is also often written as

$$\sum_{i=1}^n f(x_i^*) \Delta x$$

where $\Delta x = \frac{b-a}{n}$.

- If we choose each $x_{i,n}^* = x_{i-1} = a + (i-1)\frac{b-a}{n}$ to be the left hand end point of the i^{th} interval, $[x_{i-1}, x_i]$, we get the approximation

$$\sum_{i=1}^n f\left(a + (i-1)\frac{b-a}{n}\right) \frac{b-a}{n}$$

which is called the “left Riemann sum approximation to $\int_a^b f(x) dx$ with n subintervals”. This is the approximation used in Example 1.1.1.

- In the same way, if we choose $x_{i,n}^* = x_i = a + i\frac{b-a}{n}$ we obtain the approximation

$$\sum_{i=1}^n f\left(a + i\frac{b-a}{n}\right) \frac{b-a}{n}$$

which is called the “right Riemann sum approximation to $\int_a^b f(x) dx$ with n subintervals”. The word “right” signifies that, on each subinterval $[x_{i-1}, x_i]$ we approximate f by its value at the right-hand end-point, $x_i = a + i\frac{b-a}{n}$, of the subinterval.

- A third commonly used approximation is

$$\sum_{i=1}^n f\left(a + \left(i - \frac{1}{2}\right)\frac{b-a}{n}\right) \frac{b-a}{n}$$

which is called the “midpoint Riemann sum approximation to $\int_a^b f(x) dx$ with n subintervals”. The word “midpoint” signifies that, on each subinterval $[x_{i-1}, x_i]$ we approximate f by its value at the midpoint, $\frac{x_{i-1} + x_i}{2} = a + \left(i - \frac{1}{2}\right)\frac{b-a}{n}$, of the subinterval.

In order to compute a definite integral using Riemann sums we need to be able to compute the limit of the sum as the number of summands goes to infinity. This approach is not always feasible and we will soon arrive at other means of computing definite integrals based on antiderivatives. However, Riemann sums also provide us with a good means of approximating definite integrals — if we take n to be a large, but finite, integer, then the corresponding Riemann sum can be a good approximation of the definite integral. Under certain circumstances this can be strengthened to give rigorous bounds on the integral. Let us revisit Example 1.1.1.

✓ Example 1.1.12 Upper and lower bounds on area

Let's say we are again interested in the integral $\int_0^1 e^x dx$. We can follow the same procedure as we used previously to construct Riemann sum approximations. However since the integrand $f(x) = e^x$ is an increasing function, we can make our approximations into upper and lower bounds without much extra work.

More precisely, we approximate $f(x)$ on each subinterval $x_{i-1} \leq x \leq x_i$

- by its smallest value on the subinterval, namely $f(x_{i-1})$, when we compute the left Riemann sum approximation and
- by its largest value on the subinterval, namely $f(x_i)$, when we compute the right Riemann sum approximation.

This is illustrated in the two figures below. The shaded region in the left hand figure is the left Riemann sum approximation and the shaded region in the right hand figure is the right Riemann sum approximation.



We can see that exactly because $f(x)$ is increasing, the left Riemann sum describes an area smaller than the definite integral while the right Riemann sum gives an area larger¹⁷ than the integral.

When we approximate the integral $\int_0^1 e^x dx$ using n subintervals, then, on interval number i ,

- x runs from $\frac{i-1}{n}$ to $\frac{i}{n}$ and
- $y = e^x$ runs from $e^{\frac{i-1}{n}}$, when x is at the left hand end point of the interval, to $e^{\frac{i}{n}}$, when x is at the right hand end point of the interval.

Consequently, the left Riemann sum approximation to $\int_0^1 e^x dx$ is $\sum_{i=1}^n e^{\frac{i-1}{n}} \frac{1}{n}$ and the right Riemann sum approximation is $\sum_{i=1}^n e^{\frac{i}{n}} \cdot \frac{1}{n}$. So

$$\sum_{i=1}^n e^{\frac{(i-1)}{n}} \frac{1}{n} \leq \int_0^1 e^x dx \leq \sum_{i=1}^n e^{\frac{i}{n}} \cdot \frac{1}{n}$$

Thus $L_n = \sum_{i=1}^n e^{\frac{(i-1)}{n}} \frac{1}{n}$, which for any n can be evaluated by computer, is a lower bound on the exact value of $\int_0^1 e^x dx$ and $R_n = \sum_{i=1}^n e^{\frac{i}{n}} \frac{1}{n}$, which for any n can also be evaluated by computer, is an upper bound on the exact value of $\int_0^1 e^x dx$. For example, when $n = 1000$, $L_n = 1.7174$ and $R_n = 1.7191$ (both to four decimal places) so that, again to four decimal places,

$$1.7174 \leq \int_0^1 e^x dx \leq 1.7191$$

Recall that the exact value is $e - 1 = 1.718281828 \dots$

Using Known Areas to Evaluate Integrals

One of the main aims of this course is to build up general machinery for computing definite integrals (as well as interpreting and applying them). We shall start on this soon, but not quite yet. We have already seen one concrete, if laborious, method for computing definite integrals — taking limits of Riemann sums as we did in Example 1.1.1. A second method, which will work for some special integrands, works by interpreting the definite integral as “signed area”. This approach will work nicely when the area under the curve decomposes into simple geometric shapes like triangles, rectangles and circles. Here are some examples of this second method.

✓ Example 1.1.13 A very simple integral and a very simple area

The integral $\int_a^b 1 dx$ (which is also written as just $\int_a^b dx$) is the area of the shaded rectangle (of width $b - a$ and height 1) in the figure on the right below. So

$$\int_a^b dx = (b - a) \times (1) = b - a$$



✓ Example 1.1.14 Another simple integral

Let $b > 0$. The integral $\int_0^b x dx$ is the area of the shaded triangle (of base b and of height b) in the figure on the right below. So

$$\int_0^b x dx = \frac{1}{2} b \times b = \frac{b^2}{2}$$



The integral $\int_{-b}^0 x dx$ is the signed area of the shaded triangle (again of base b and of height b) in the figure on the right below. So

$$\int_{-b}^0 x dx = -\frac{b^2}{2}$$



Notice that it is very easy to extend this example to the integral $\int_0^b cx dx$ for any real numbers $b, c > 0$ and find

$$\int_0^b cx dx = \frac{c}{2} b^2.$$

✓ Example 1.1.15 Evaluating $\int_{-1}^1 (1 - |x|) dx$

In this example, we shall evaluate $\int_{-1}^1 (1 - |x|) dx$. Recall that

$$|x| = \begin{cases} -x & \text{if } x \leq 0 \\ x & \text{if } x \geq 0 \end{cases}$$

so that

$$1 - |x| = \begin{cases} 1 + x & \text{if } x \leq 0 \\ 1 - x & \text{if } x \geq 0 \end{cases}$$

To picture the geometric figure whose area the integral represents observe that

- at the left hand end of the domain of integration $x = -1$ and the integrand $1 - |x| = 1 - |-1| = 1 - 1 = 0$ and
- as x increases from -1 towards 0 , the integrand $1 - |x| = 1 + x$ increases linearly, until
- when x hits 0 the integrand hits $1 - |x| = 1 - |0| = 1$ and then
- as x increases from 0 , the integrand $1 - |x| = 1 - x$ decreases linearly, until
- when x hits $+1$, the right hand end of the domain of integration, the integrand hits $1 - |x| = 1 - |1| = 0$.

So the integral $\int_{-1}^1 (1 - |x|) dx$ is the area of the shaded triangle (of base 2 and of height 1) in the figure on the right below and

$$\int_{-1}^1 (1 - |x|) dx = \frac{1}{2} \times 2 \times 1 = 1$$



✓ Example 1.1.16. Evaluating $\int_0^1 \sqrt{1-x^2} dx$

The integral $\int_0^1 \sqrt{1-x^2} dx$ has integrand $f(x) = \sqrt{1-x^2}$. So it represents the area under $y = \sqrt{1-x^2}$ with x running from 0 to 1 . But we may rewrite

$$y = \sqrt{1-x^2} \quad \text{as} \quad x^2 + y^2 = 1, y \geq 0$$

But this is the (implicit) equation for a circle — the extra condition that $y \geq 0$ makes it the equation for the semi-circle centered at the origin with radius 1 lying on and above the x -axis. Thus the integral represents the area of the quarter circle of radius 1 , as shown in the figure on the right below. So

$$\int_0^1 \sqrt{1-x^2} dx = \frac{1}{4} \pi (1)^2 = \frac{\pi}{4}$$



This next one is a little trickier and relies on us knowing the symmetries of the sine function.

✓ Example 1.1.17 Integrating sine

The integral $\int_{-\pi}^{\pi} \sin x dx$ is the signed area of the shaded region in the figure on the right below. It naturally splits into two regions, one on either side of the y -axis. We don't know the formula for the area of either of these regions (yet), however the two regions are very nearly the same. In fact, the part of the shaded region below the x -axis is exactly the reflection, in the x -axis, of the part of the shaded region above the x -axis. So the signed area of part of the shaded region below the x -axis is the negative of the signed area of part of the shaded region above the x -axis and

$$\int_{-\pi}^{\pi} \sin x dx = 0$$



Another Interpretation for Definite Integrals

So far, we have only a single interpretation¹⁸ for definite integrals — namely areas under graphs. In the following example, we develop a second interpretation.

✓ Example 1.1.18 A moving particle

Suppose that a particle is moving along the x -axis and suppose that at time t its velocity is $v(t)$ (with $v(t) > 0$ indicating rightward motion and $v(t) < 0$ indicating leftward motion). What is the change in its x -coordinate between time a and time $b > a$?

We'll work this out using a procedure similar to our definition of the integral. First pick a natural number n and divide the time interval from a to b into n equal subintervals, each of width $\frac{b-a}{n}$. We are working our way towards a Riemann sum (as we have done several times above) and so we will eventually take the limit $n \rightarrow \infty$.

- The first time interval runs from a to $a + \frac{b-a}{n}$. If we think of n as some large number, the width of this interval, $\frac{b-a}{n}$ is very small and over this time interval, the velocity does not change very much. Hence we can approximate the velocity over the first subinterval as being essentially constant at its value at the start of the time interval — $v(a)$. Over the

subinterval the x -coordinate changes by velocity times time, namely $v(a) \cdot \frac{b-a}{n}$.

- Similarly, the second interval runs from time $a + \frac{b-a}{n}$ to time $a + 2\frac{b-a}{n}$. Again, we can assume that the velocity does not change very much and so we can approximate the velocity as being essentially constant at its value at the start of the subinterval — namely $v\left(a + \frac{b-a}{n}\right)$. So during the second subinterval the particle's x -coordinate changes by approximately $v\left(a + \frac{b-a}{n}\right) \frac{b-a}{n}$.
- In general, time subinterval number i runs from $a + (i-1)\frac{b-a}{n}$ to $a + i\frac{b-a}{n}$ and during this subinterval the particle's x -coordinate changes, essentially, by

$$, v\left(a + (i-1)\frac{b-a}{n}\right) \frac{b-a}{n}.$$

So the net change in x -coordinate from time a to time b is approximately

$$\begin{aligned} & v(a) \frac{b-a}{n} + v\left(a + \frac{b-a}{n}\right) \frac{b-a}{n} + \cdots + v\left(a + (i-1)\frac{b-a}{n}\right) \frac{b-a}{n} + \cdots \\ & + v\left(a + (n-1)\frac{b-a}{n}\right) \frac{b-a}{n} \\ & = \sum_{i=1}^n v\left(a + (i-1)\frac{b-a}{n}\right) \frac{b-a}{n} \end{aligned}$$

This exactly the left Riemann sum approximation to the integral of v from a to b with n subintervals. The limit as $n \rightarrow \infty$ is exactly the definite integral $\int_a^b v(t) dt$. Following tradition, we have called the (dummy) integration variable t rather than x to remind us that it is time that is running from a to b .

The conclusion of the above discussion is that if a particle is moving along the x -axis and its x -coordinate and velocity at time t are $x(t)$ and $v(t)$, respectively, then, for all $b > a$,

$$x(b) - x(a) = \int_a^b v(t) dt.$$

Optional — careful definition of the integral

In this optional section we give a more mathematically rigorous definition of the definite integral $\int_a^b f(x) dx$. Some textbooks use a sneakier, but equivalent, definition. The integral will be defined as the limit of a family of approximations to the area between the graph of $y = f(x)$ and the x -axis, with x running from a to b . We will then show conditions under which this limit is guaranteed to exist. We should state up front that these conditions are more restrictive than is strictly necessary — this is done so as to keep the proof accessible.

The family of approximations needed is slightly more general than that used to define Riemann sums in the previous sections, though it is quite similar. The main difference is that we do not require that all the subintervals have the same size.

- We start by selecting a positive integer n . As was the case previously, this will be the number of subintervals used in the approximation and eventually we will take the limit as $n \rightarrow \infty$.
- Now subdivide the interval from a to b into n subintervals by selecting $n+1$ values of x that obey

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

The subinterval number i runs from x_{i-1} to x_i . This formulation does not require the subintervals to have the same size. However we will eventually require that the widths of the subintervals shrink towards zero as $n \rightarrow \infty$.

- Then for each subinterval we select a value of x in that interval. That is, for $i = 1, 2, \dots, n$, choose x_i^* satisfying $x_{i-1} \leq x_i^* \leq x_i$. We will use these values of x to help approximate $f(x)$ on each subinterval.
- The area between the graph of $y = f(x)$ and the x -axis, with x running

from x_{i-1} to x_i , i.e. the contribution, $\int_{x_{i-1}}^{x_i} f(x) dx$, from interval number i to the integral, is approximated by the area of a rectangle. The rectangle has width $x_i - x_{i-1}$ and height $f(x_i^*)$.

- Thus the approximation to the integral, using all n subintervals, is

$$\int_a^b f(x) dx \approx f(x_1^*)[x_1 - x_0] + f(x_2^*)[x_2 - x_1] + \cdots + f(x_n^*)[x_n - x_{n-1}]$$

- Of course every different choice of n and x_1, x_2, \dots, x_{n-1} and $x_1^*, x_2^*, \dots, x_n^*$ gives a different approximation. So to simplify the discussion that follows, let us denote a particular choice of all these numbers by \mathbb{P} :

$$\mathbb{P} = (n, x_1, x_2, \dots, x_{n-1}, x_1^*, x_2^*, \dots, x_n^*).$$

Similarly let us denote the resulting approximation by $\backslash \mathbf{cI}(\mathbb{P})$:

$$\backslash \mathbf{cI}(\mathbb{P}) = f(x_1^*)[x_1 - x_0] + f(x_2^*)[x_2 - x_1] + \cdots + f(x_n^*)[x_n - x_{n-1}]$$

- We claim that, for any reasonable¹⁹ function $f(x)$, if you take any reasonable²⁰ sequence of these approximations you always get the exactly the same limiting value. We define $\int_a^b f(x) dx$ to be this limiting value.
- Let's be more precise. We can take the limit of these approximations in two equivalent ways. Above we did this by taking the number of subintervals n to infinity. When we did this, the width of all the subintervals went to zero. With the formulation we are now using, simply taking the number of subintervals to be very large does not imply that they will all shrink in size. We could have one very large subinterval and a large number of tiny ones. Thus we take the limit we need by taking the width of the subintervals to zero. So for any choice \mathbb{P} , we define

$$M(\mathbb{P}) = \max \{x_1 - x_0, x_2 - x_1, \dots, x_n - x_{n-1}\}$$

that is the maximum width of the subintervals used in the approximation determined by \mathbb{P} . By forcing the maximum width to go to zero, the widths of all the subintervals go to zero.

- We then define the definite integral as the limit

$$\int_a^b f(x) dx = \lim_{M(\mathbb{P}) \rightarrow 0} \backslash \mathbf{cI}(\mathbb{P}).$$

Of course, one is now left with the question of determining when the above limit exists. A proof of the very general conditions which guarantee existence of this limit is beyond the scope of this course, so we instead give a weaker result (with stronger conditions) which is far easier to prove.

For the rest of this section, assume

- that $f(x)$ is continuous for $a \leq x \leq b$,
- that $f(x)$ is differentiable for $a < x < b$, and
- that $f'(x)$ is bounded — ie $|f'(x)| \leq F$ for some constant F .

We will now show that, under these hypotheses, as $M(\mathbb{P})$ approaches zero, $\backslash \mathbf{cI}(\mathbb{P})$ always approaches the area, A , between the graph of $y = f(x)$ and the x -axis, with x running from a to b .

These assumptions are chosen to make the argument particularly transparent. With a little more work one can weaken the hypotheses considerably. We are cheating a little by implicitly assuming that the area A exists. In fact, one can adjust the argument below to remove this implicit assumption.

- Consider A_j , the part of the area coming from $x_{j-1} \leq x \leq x_j$.

We have approximated this area by $f(x_j^*)[x_j - x_{j-1}]$ (see figure left).

- Let $f(\bar{x}_j)$ and $f(\underline{x}_j)$ be the largest and smallest values²¹ of $f(x)$ for $x_{j-1} \leq x \leq x_j$. Then the true area is bounded by

$$f(\underline{x}_j)[x_j - x_{j-1}] \leq A_j \leq f(\bar{x}_j)[x_j - x_{j-1}].$$

(see figure right).

- Now since $f(\underline{x}_j) \leq f(x_j^*) \leq f(\bar{x}_j)$, we also know that

$$f(\underline{x}_j)[x_j - x_{j-1}] \leq f(x_j^*)[x_j - x_{j-1}] \leq f(\bar{x}_j)[x_j - x_{j-1}].$$

- So both the true area, A_j , and our approximation of that area $f(x_j^*)[x_j - x_{j-1}]$ have to lie between $f(\bar{x}_j)[x_j - x_{j-1}]$ and $f(\underline{x}_j)[x_j - x_{j-1}]$. Combining these bounds we have that the difference between the true area and our approximation of that area is bounded by

$$|A_j - f(x_j^*)[x_j - x_{j-1}]| \leq [f(\bar{x}_j) - f(\underline{x}_j)] \cdot [x_j - x_{j-1}].$$

(To see this think about the smallest the true area can be and the largest our approximation can be and vice versa.)

- Now since our function, $f(x)$ is differentiable we can apply one of the main theorems we learned in CLP-1 — the Mean Value Theorem²². The MVT implies that there exists a c between \underline{x}_j and \bar{x}_j such that

$$f(\bar{x}_j) - f(\underline{x}_j) = f'(c) \cdot [\bar{x}_j - \underline{x}_j]$$

- By the assumption that $|f'(x)| \leq F$ for all x and the fact that \underline{x}_j and \bar{x}_j must both be between x_{j-1} and x_j

$$|f(\bar{x}_j) - f(\underline{x}_j)| \leq F \cdot |\bar{x}_j - \underline{x}_j| \leq F \cdot [x_j - x_{j-1}]$$

Hence the error in this part of our approximation obeys

$$|A_j - f(x_j^*)[x_j - x_{j-1}]| \leq F \cdot [x_j - x_{j-1}]^2.$$

- That was just the error in approximating A_j . Now we bound the total error by combining the errors from approximating on all the subintervals. This gives

$$\begin{aligned} |A - \text{cI}(\mathbb{P})| &= \left| \sum_{j=1}^n A_j - \sum_{j=1}^n f(x_j^*)[x_j - x_{j-1}] \right| \\ &= \left| \sum_{j=1}^n (A_j - f(x_j^*)[x_j - x_{j-1}]) \right| && \text{triangle inequality} \\ &\leq \sum_{j=1}^n |A_j - f(x_j^*)[x_j - x_{j-1}]| \\ &\leq \sum_{j=1}^n F \cdot [x_j - x_{j-1}]^2 && \text{from above} \end{aligned}$$

Now do something a little sneaky. Replace one of these factors of $[x_j - x_{j-1}]$ (which is just the width of the j^{th} subinterval) by the maximum width of the subintervals:

$$\begin{aligned} &\leq \sum_{j=1}^n F \cdot M(\mathbb{P}) \cdot [x_j - x_{j-1}] && F \text{ and } M(\mathbb{P}) \text{ are constant} \\ &\leq F \cdot M(\mathbb{P}) \cdot \sum_{j=1}^n [x_j - x_{j-1}] && \text{sum is total width} \\ &= F \cdot M(\mathbb{P}) \cdot (b - a). \end{aligned}$$

- Since a , b and F are fixed, this tends to zero as the maximum rectangle width $M(\mathbb{P})$ tends to zero.

Thus, we have proven

Theorem 1.1.19

Assume that $f(x)$ is continuous for $a \leq x \leq b$, and is differentiable for all $a < x < b$ with $|f'(x)| \leq F$, for some constant F . Then, as the maximum rectangle width $M(\mathbb{P})$ tends to zero, $\text{cI}(\mathbb{P})$ always converges to A , the area between the graph of $y = f(x)$ and the x -axis, with x running from a to b .

Exercises

Stage 1

For Questions 1 through 5, we want you to develop an understanding of the model we are using to define an integral: we approximate the area under a curve by bounding it between rectangles. Later, we will learn more sophisticated methods of integration, but they are all based on this simple concept.

In Questions 6 through 10, we practice using sigma notation. There are many ways to write a given sum in sigma notation. You can practice finding several, and deciding which looks the clearest.

Questions 11 through 15 are meant to give you practice interpreting the formulas in Definition 1.1.11. The formulas might look complicated at first, but if you understand what each piece means, they are easy to learn.

? 1

Give a range of possible values for the shaded area in the picture below.



? 1

Give a range of possible values for the shaded area in the picture below.



? 3

Using rectangles, find a lower and upper bound for $\int_1^3 \frac{1}{2^x} dx$ that differ by at most 0.2 square units.



? 4

Let $f(x)$ be a function that is *decreasing* from $x = 0$ to $x = 5$. Which Riemann sum approximation of $\int_0^5 f(x) dx$ is the largest--left, right, or midpoint?

? 5

Give an example of a function $f(x)$, an interval $[a, b]$, and a number n such that the midpoint Riemann sum of $f(x)$ over $[a, b]$ using n intervals is *larger than* both the left and right Riemann sums of $f(x)$ over $[a, b]$ using n intervals.

? 6

Express the following sums in sigma notation:

1. $3 + 4 + 5 + 6 + 7$
2. $6 + 8 + 10 + 12 + 14$
3. $7 + 9 + 11 + 13 + 15$
4. $1 + 3 + 5 + 7 + 9 + 11 + 13 + 15$

? 1

Express the following sums in sigma notation:

1. $\frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \frac{1}{81}$
2. $\frac{2}{3} + \frac{2}{9} + \frac{2}{27} + \frac{2}{81}$
3. $-\frac{2}{3} + \frac{2}{9} - \frac{2}{27} + \frac{2}{81}$
4. $\frac{2}{3} - \frac{2}{9} + \frac{2}{27} - \frac{2}{81}$

? 8

Express the following sums in sigma notation:

1. $\frac{1}{3} + \frac{1}{3} + \frac{5}{27} + \frac{7}{81} + \frac{9}{243}$

2. $\frac{1}{5} + \frac{1}{11} + \frac{1}{29} + \frac{1}{83} + \frac{1}{245}$

3. $1000 + 200 + 30 + 4 + \frac{1}{2} + \frac{3}{50} + \frac{7}{1000}$

? 9

Evaluate the following sums. You might want to use the formulas from Theorems 5 and 6.

1. $\sum_{i=0}^{100} \left(\frac{3}{5}\right)^i$

2. $\sum_{i=50}^{100} \left(\frac{3}{5}\right)^i$

3. $\sum_{i=1}^{10} (i^2 - 3i + 5)$

4. $\sum_{n=1}^b \left[\left(\frac{1}{e}\right)^n + en^3 \right]$, where b is some integer greater than 1.

? 10

Evaluate the following sums. You might want to use the formulas from Theorem 1.1.6.

1. $\sum_{i=50}^{100} (i - 50) + \sum_{i=0}^{50} i$

2. $\sum_{i=10}^{100} (i - 5)^3$

3. $\sum_{n=1}^{11} (-1)^n$

4. $\sum_{n=2}^{11} (-1)^{2n+1}$

? 11

In the picture below, draw in the rectangles whose (signed) area is being computed by the midpoint Riemann sum

$$\sum_{i=1}^4 \frac{b-a}{4} \cdot f\left(a + \left(i - \frac{1}{2}\right) \frac{b-a}{4}\right).$$



? 12 (*)

$\sum_{k=1}^4 f(1+k) \cdot 1$ is a left Riemann sum for a function $f(x)$ on the interval $[a, b]$ with n subintervals. Find the values of a , b and n .

? 13

Draw a picture illustrating the area given by the following Riemann sum.

$$\sum_{i=1}^3 2 \cdot (5 + 2i)^2$$

? 14

Draw a picture illustrating the area given by the following Riemann sum.

$$\sum_{i=1}^5 \frac{\pi}{20} \cdot \tan\left(\frac{\pi(i-1)}{20}\right)$$

? 15 (*)

Fill in the blanks with right, left, or midpoint; an interval; and a value of n .

1. $\sum_{k=0}^3 f(1.5 + k) \cdot 1$ is a _____ Riemann sum for f on the interval [____, ____] with $n =$ ____.

? 16

Evaluate the following integral by interpreting it as a signed area, and using geometry:

$$\int_0^5 x \, dx$$

? 17

Evaluate the following integral by interpreting it as a signed area, and using geometry:

$$\int_{-2}^5 x \, dx$$

Stage 2

Questions 26 and 27 use the formula for a geometric sum, Equation 1.1.3

Remember that a definite integral is a signed area between a curve and the x -axis. We'll spend a lot of time learning strategies for evaluating definite integrals, but we already know lots of ways to find area of geometric shapes. In Questions 28 through 33, use your knowledge of geometry to find the signed areas described by the integrals given.

? 18 (*)

Use sigma notation to write the midpoint Riemann sum for $f(x) = x^8$ on $[5, 15]$ with $n = 50$. Do not evaluate the Riemann sum.

? 19 (*)

Estimate $\int_{-1}^5 x^3 \, dx$ using three approximating rectangles and left hand end points.

? 20 (*)

Let f be a function on the whole real line. Express $\int_{-1}^7 f(x) dx$ as a limit of Riemann sums, using the right endpoints.

? 21 (*)

The value of the following limit is equal to the area below a graph of $y = f(x)$, integrated over the interval $[0, b]$:

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{4}{n} \left[\sin \left(2 + \frac{4i}{n} \right) \right]^2$$

Find $f(x)$ and b .

? 22 (*)

For a certain function $f(x)$, the following equation holds:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{k}{n^2} \sqrt{1 - \frac{k^2}{n^2}} = \int_0^1 f(x) dx$$

Find $f(x)$.

? 23 (*)

Express $\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{3}{n} e^{-i/n} \cos \left(\frac{3i}{n} \right)$ as a definite integral.

? 24 (*)

Let $R_n = \sum_{i=1}^n \frac{i e^{i/n}}{n^2}$. Express $\lim_{n \rightarrow \infty} R_n$ as a definite integral. *Do not evaluate this integral.*

? 25 (*)

Express $\lim_{n \rightarrow \infty} \left(\sum_{i=1}^n e^{-1-2i/n} \cdot \frac{2}{n} \right)$ as an integral in three different ways.

? 26

Evaluate the sum $1 + r^3 + r^6 + r^9 + \dots + r^{3n}$.

? 27

Evaluate the sum $r^5 + r^6 + r^7 + \dots + r^{100}$.

? 28 (*)

Evaluate $\int_{-1}^2 |2x| dx$.

? 29

Evaluate the following integral by interpreting it as a signed area, and using geometry:

$$\int_{-3}^5 |t - 1| dt$$

? 30

Evaluate the following integral by interpreting it as a signed area, and using geometry:

$$\int_a^b x dx$$

where $0 \leq a \leq b$.

? 31

Evaluate the following integral by interpreting it as a signed area, and using geometry:

$$\int_a^b x dx$$

where $a \leq b \leq 0$.

? 32

Evaluate the following integral by interpreting it as a signed area, and using geometry:

$$\int_0^4 \sqrt{16 - x^2} dx$$

? 33 (*)

Use elementary geometry to calculate $\int_0^3 f(x) dx$, where

$$f(x) = \begin{cases} x, & \text{if } x \leq 1, \\ 1, & \text{if } x > 1. \end{cases}$$

? 34 (*)

A car's gas pedal is applied at $t = 0$ seconds and the car accelerates continuously until $t = 2$ seconds. The car's speed at half-second intervals is given in the table below. Find the best possible **upper** estimate for the distance that the car traveled during these two seconds.

t (s)	0	0.5	1.0	1.5	2
v (m/s)	0	14	22	30	40

? 35

True or false: the answer you gave for Question 34 is definitely greater than or equal to the distance the car travelled during the two seconds in question.

? 36

An airplane's speed at one-hour intervals is given in the table below. Approximate the distance travelled by the airplane from noon to 4pm three ways using a midpoint Riemann sum.

time	12:00 pm	1:00 pm	2:00 pm	3:00 pm	4:00 pm
speed (km/hr)	800	700	850	900	750

Stage 3

? 37 (*)

(a) Express

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{2}{n} \sqrt{4 - \left(-2 + \frac{2i}{n}\right)^2}$$

as a definite integral.

(b) Evaluate the integral of part (a).

? 38 (*)

Consider the integral:

$$\int_0^3 (7 + x^3) dx. \quad (*)$$

1. Approximate this integral using the left Riemann sum with $n = 3$ intervals.
2. Write down the expression for the right Riemann sum with n intervals and calculate the sum. Now take the limit $n \rightarrow \infty$ in your expression for the Riemann sum, to evaluate the integral (*) exactly.

You may use the identity

$$\sum_{i=1}^n i^3 = \frac{n^4 + 2n^3 + n^2}{4}$$

? 39 (*)

Using a limit of right-endpoint Riemann sums, evaluate $\int_2^4 x^2 dx$. You may use the formulas $\sum_{i=1}^n i = \frac{n(n+1)}{2}$ and

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}.$$

? 40 (*)

Find $\int_0^2 (x^3 + x) dx$ using the definition of the definite integral. You may use the summation formulas $\sum_{i=1}^n i^3 = \frac{n^4 + 2n^3 + n^2}{4}$

and $\sum_{i=1}^n i = \frac{n^2 + n}{2}.$

? 41 (*)

Using a limit of right-endpoint Riemann sums, evaluate $\int_1^4 (2x - 1) dx$. Do not use anti-differentiation, except to check your answer. ²³You'll learn about this method starting in Section 1.3. You can also check this answer using geometry. You may use the formula $\sum_{i=1}^n i = \frac{n(n+1)}{2}$.

? 42

Give a function $f(x)$ that has the following expression as a right Riemann sum when $n = 10$, $\Delta(x) = 10$ and $a = -5$:

$$\sum_{i=1}^{10} 3(7 + 2i)^2 \sin(4i).$$

? 43

Using the method of Example 1.1.2, evaluate

$$\int_0^1 2^x dx$$

? 44

1. Using the method of Example 1.1.2, evaluate

$$\int_a^b 10^x dx$$

2.

Using your answer from above, make a guess for

$$\int_a^b c^x dx$$

where c is a positive constant. Does this agree with Question 43?

? 45

Evaluate $\int_0^a \sqrt{1-x^2} dx$ using geometry, if $0 \leq a \leq 1$.

? 46

Suppose $f(x)$ is a positive, decreasing function from $x = a$ to $x = b$. You give an upper and lower bound on the area under the curve $y = f(x)$ using n rectangles and a left and right Riemann sum, respectively, as in the picture below.



1. What is the difference between the lower bound and the upper bound? (That is, if we subtract the smaller estimate from the larger estimate, what do we get?) Give your answer in terms of f , a , b , and n .
2. If you want to approximate the area under the curve to within 0.01 square units using this method, how many rectangles should you use? That is, what should n be?

Let $f(x)$ be a linear function, let $a < b$ be integers, and let n be a whole number. True or false: if we average the left and right Riemann sums for $\int_a^b f(x) dx$ using n rectangles, we get the same value as the midpoint Riemann sum using n rectangles.

1. This should remind the reader of the approach taken to compute the slope of a tangent line way way back at the start of differential calculus.
2. Approximating the area in this way leads to a definition of integration that is called Riemann integration. This is the most commonly used approach to integration. However we could also approximate the area by using long thin horizontal strips. This leads to a definition of integration that is called Lebesgue integration. We will not be covering Lebesgue integration in these notes.
3. If we want to be more careful here, we should construct two approximations, one that is always a little smaller than the desired area and one that is a little larger. We can then take a limit using the Squeeze Theorem and arrive at the exact area. More on this later.
4. Notice that since e^x is an increasing function, this choice of heights means that each of our rectangles is smaller than the strip it came from.
5. We haven't proved that this will give us the exact area, but it should be clear that taking this limit will give us a lower bound on the area. To complete things rigorously we also need an upper bound and the squeeze theorem. We do this in the next optional subsection.
6. If you do not recall L'Hôpital's rule and indeterminate forms then we recommend you skim over your differential calculus notes on the topic.
7. Say if you don't recall l'Hôpital's rule and have not had time to revise it.
8. To hyphenate or not to hyphenate: “non-zero” or “nonzero”? The authors took our lead from [here](#) and also [here](#).
9. Recall that if we have 3 functions $f(x), g(x), h(x)$ that satisfy $f(x) \leq g(x) \leq h(x)$ and we know that $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} h(x) = L$ exists and is finite, then the squeeze theorem tells us that $\lim_{x \rightarrow a} g(x) = L$.
10. Some careful addition shows it is $\frac{46181}{176400}$.
11. Since all the sums are finite, this isn't too hard. More care must be taken when the sums involve an infinite number of terms. We will examine this in Chapter 3.
12. Of course, any finite sum can be computed exactly — just sum together the terms. What we mean by “computed exactly” in this context, is that we can rewrite the sum as a simple, and easily evaluated, formula involving the terminals of the sum. For example $\sum_{k=m}^n r^k = \frac{r^{n+1} - r^m}{r - 1}$ provided $r \neq 1$. No matter what finite integers we choose for m and n , we can quickly compute the sum in just a few arithmetic operations. On the other hand, the sums, $\sum_{k=m}^n \frac{1}{k}$ and $\sum_{k=m}^n \frac{1}{k^2}$, cannot be expressed in such clean formulas (though you can rewrite them quite cleanly using integrals). To explain more clearly we would need to go into a more detailed and careful discussion that is beyond the scope of this course.
13. Generating functions are frequently used in mathematics to analyse sequences and series, but are beyond the scope of the course. The interested reader should take a look at “Generatingfunctionology” by Herb Wilf. It is an excellent book and is also free to download.
14. a and b are also called the bounds of integration.
15. We'll eventually allow a and b to be any two real numbers, not even requiring $a < b$. But it is easier to start off assuming $a < b$, and that's what we'll do.
16. Bernhard Riemann was a 19th century German mathematician who made extremely important contributions to many different areas of mathematics — far too many to list here. Arguably two of the most important (after Riemann sums) are now called Riemann surfaces and the Riemann hypothesis (he didn't name them after himself).
17. When a function is decreasing the situation is reversed — the left Riemann sum is always larger than the integral while the right Riemann sum is smaller than the integral. For more general functions that both increase and decrease it is perhaps easiest to study each increasing (or decreasing) interval separately.
18. If this were the only interpretation then integrals would be a nice mathematical curiosity and unlikely to be the core topic of a large first year mathematics course.
19. We'll be more precise about what “reasonable” means shortly.
20. Again, we'll explain this “reasonable” shortly

21. Here we are using the extreme value theorem — its proof is beyond the scope of this course. The theorem says that any continuous function on a closed interval must attain a minimum and maximum at least once. In this situation this implies that for any continuous function $f(x)$, there are $x_{j-1} \leq \bar{x}_j, \underline{x}_j \leq x_j$ such that $f(\underline{x}_j) \leq f(x) \leq f(\bar{x}_j)$ for all $x_{j-1} \leq x \leq x_j$.
22. Recall that the mean value theorem states that for a function continuous on $[a, b]$ and differentiable on (a, b) , there exists a number c between a and b so that $f'(c) = \frac{f(b)-f(a)}{b-a}$.

This page titled [1.1: Definition of the Integral](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.2: Basic properties of the definite integral

When we studied limits and derivatives, we developed methods for taking limits or derivatives of “complicated functions” like $f(x) = x^2 + \sin(x)$ by understanding how limits and derivatives interact with basic arithmetic operations like addition and subtraction. This allowed us to reduce the problem into one of computing derivatives of simpler functions like x^2 and $\sin(x)$. Along the way we established simple rules such as

$$\lim_{x \rightarrow a} (f(x) + g(x)) = \lim_{x \rightarrow a} f(x) + \lim_{x \rightarrow a} g(x) \quad \text{and} \quad \frac{d}{dx} (f(x) + g(x)) = \frac{df}{dx} + \frac{dg}{dx}$$

Some of these rules have very natural analogues for integrals and we discuss them below. Unfortunately the analogous rules for integrals of products of functions or integrals of compositions of functions are more complicated than those for limits or derivatives. We discuss those rules at length in subsequent sections. For now let us consider some of the simpler rules of the arithmetic of integrals.

Theorem 1.2.1 Arithmetic of Integration

Let a, b and A, B, C be real numbers. Let the functions $f(x)$ and $g(x)$ be integrable on an interval that contains a and b . Then

$$(a) \quad \int_a^b (f(x) + g(x)) \, dx = \int_a^b f(x) \, dx + \int_a^b g(x) \, dx$$

$$(b) \quad \int_a^b (f(x) - g(x)) \, dx = \int_a^b f(x) \, dx - \int_a^b g(x) \, dx$$

$$(c) \quad \int_a^b C f(x) \, dx = C \cdot \int_a^b f(x) \, dx$$

Combining these three rules we have

$$(d) \quad \int_a^b (A f(x) + B g(x)) \, dx = A \int_a^b f(x) \, dx + B \int_a^b g(x) \, dx$$

That is, integrals depend linearly on the integrand.

$$(e) \quad \int_a^b dx = \int_a^b 1 \cdot dx = b - a$$

It is not too hard to prove this result from the definition of the definite integral. Additionally we only really need to prove (d) and (e) since

- (a) follows from (d) by setting $A = B = 1$,
- (b) follows from (d) by setting $A = 1, B = -1$, and
- (c) follows from (d) by setting $A = C, B = 0$.

Proof

As noted above, it suffices for us to prove (d) and (e). Since (e) is easier, we will start with that. It is also a good warm-up for (d).

- The definite integral in (e), $\int_a^b 1 \, dx$, can be interpreted geometrically as the area of the rectangle with height 1 running from $x = a$ to $x = b$; this area is clearly $b - a$. We can also prove this formula from the definition of the integral (Definition 1.1.9):

$$\begin{aligned}
 \int_a^b dx &= \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_{i,n}^*) \frac{b-a}{n} && \text{by definition} \\
 &= \lim_{n \rightarrow \infty} \sum_{i=1}^n 1 \frac{b-a}{n} && \text{since } f(x) = 1 \\
 &= \lim_{n \rightarrow \infty} (b-a) \sum_{i=1}^n \frac{1}{n} && \text{since } a, b \text{ are constants} \\
 &= \lim_{n \rightarrow \infty} (b-a) \\
 &= b-a
 \end{aligned}$$

as required.

- To prove (d) let us start by defining $h(x) = Af(x) + Bg(x)$ and then we need to express the integral of $h(x)$ in terms of those of $f(x)$ and $g(x)$. We use Definition 1.1.9 and some algebraic manipulations¹ to arrive at the result.

$$\begin{aligned}
 \int_a^b h(x) dx &= \sum_{i=1}^n h(x_{i,n}^*) \cdot \frac{b-a}{n} \\
 &\quad \text{by Definition 1.1.9} \\
 &= \sum_{i=1}^n \left(Af(x_{i,n}^*) + Bg(x_{i,n}^*) \right) \cdot \frac{b-a}{n} \\
 &= \sum_{i=1}^n \left(Af(x_{i,n}^*) \cdot \frac{b-a}{n} + Bg(x_{i,n}^*) \cdot \frac{b-a}{n} \right) \\
 &= \left(\sum_{i=1}^n Af(x_{i,n}^*) \cdot \frac{b-a}{n} \right) + \left(\sum_{i=1}^n Bg(x_{i,n}^*) \cdot \frac{b-a}{n} \right) \\
 &\quad \text{by Theorem 1.1.5(b)} \\
 &= A \left(\sum_{i=1}^n f(x_{i,n}^*) \cdot \frac{b-a}{n} \right) + B \left(\sum_{i=1}^n g(x_{i,n}^*) \cdot \frac{b-a}{n} \right) \\
 &\quad \text{by Theorem 1.1.5(a)} \\
 &= A \int_a^b f(x) dx + B \int_a^b g(x) dx \\
 &\quad \text{by Definition 1.1.9}
 \end{aligned}$$

as required.

Using this Theorem we can integrate sums, differences and constant multiples of functions we know how to integrate. For example:

✓ Example 1.2.2 The integral of a sum

In Example 1.1.1 we saw that $\int_0^1 e^x dx = e - 1$. So

$$\begin{aligned}
 \int_0^1 (e^x + 7) dx &= \int_0^1 e^x dx + 7 \int_0^1 1 dx \\
 &\quad \text{by Theorem 1.2.1(d) with } A = 1, f(x) = e^x, B = 7, g(x) = 1 \\
 &= (e - 1) + 7 \times (1 - 0) \\
 &\quad \text{by Example 1.1.1 and Theorem 1.2.1(e)} \\
 &= e + 6
 \end{aligned}$$

When we gave the formal definition of $\int_a^b f(x) dx$ in Definition 1.1.9 we explained that the integral could be interpreted as the signed area between the curve $y = f(x)$ and the x -axis on the interval $[a, b]$. In order for this interpretation to make sense we required that $a < b$, and though we remarked that the integral makes sense when $a > b$ we did not explain any further. Thankfully there is an easy way to express the integral $\int_a^b f(x) dx$ in terms of $\int_b^a f(x) dx$ — making it always possible to write an integral so

the lower limit of integration is less than the upper limit of integration. Theorem 1.2.3, below, tells us that, for example, $\int_7^3 e^x dx = -\int_3^7 e^x dx$. The same theorem also provides us with two other simple manipulations of the limits of integration.

Theorem 1.2.3 Arithmetic for the Domain of Integration

Let a, b, c be real numbers. Let the function $f(x)$ be integrable on an interval that contains a, b and c . Then

$$\begin{aligned} \text{(a)} \quad & \int_a^a f(x) dx = 0 \\ \text{(b)} \quad & \int_b^a f(x) dx = -\int_a^b f(x) dx \\ \text{(c)} \quad & \int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx \end{aligned}$$

The proof of this statement is not too difficult.

Proof

Let us prove the statements in order.

- Consider the definition of the definite integral

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_{i,n}^*) \cdot \frac{b-a}{n}$$

If we now substitute $b = a$ in this expression we have

$$\begin{aligned} \int_a^a f(x) dx &= \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_{i,n}^*) \cdot \underbrace{\frac{a-a}{n}}_{=0} \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \underbrace{f(x_{i,n}^*)}_{=0} \cdot 0 \\ &= \lim_{n \rightarrow \infty} 0 \\ &= 0 \end{aligned}$$

as required.

- Consider now the definite integral $\int_a^b f(x) dx$. We will sneak up on the proof by first examining Riemann sum approximations to both this and $\int_b^a f(x) dx$. The midpoint Riemann sum approximation to $\int_a^b f(x) dx$ with 4 subintervals (so that each subinterval has width $\frac{b-a}{4}$) is

$$\begin{aligned} & \left\{ f\left(a + \frac{1}{2} \frac{b-a}{4}\right) + f\left(a + \frac{3}{2} \frac{b-a}{4}\right) + f\left(a + \frac{5}{2} \frac{b-a}{4}\right) + f\left(a + \frac{7}{2} \frac{b-a}{4}\right) \right\} \cdot \frac{b-a}{4} \\ &= \left\{ f\left(\frac{7}{8}a + \frac{1}{8}b\right) + f\left(\frac{5}{8}a + \frac{3}{8}b\right) + f\left(\frac{3}{8}a + \frac{5}{8}b\right) + f\left(\frac{1}{8}a + \frac{7}{8}b\right) \right\} \cdot \frac{b-a}{4} \end{aligned}$$

Now we do the same for $\int_b^a f(x) dx$ with 4 subintervals. Note that b is now the lower limit on the integral and a is now the upper limit on the integral. This is likely to cause confusion when we write out the Riemann sum, so we'll temporarily rename b to A and a to B . The midpoint Riemann sum approximation to $\int_A^B f(x) dx$ with 4 subintervals is

$$\begin{aligned} & \left\{ f\left(A + \frac{1}{2} \frac{B-A}{4}\right) + f\left(A + \frac{3}{2} \frac{B-A}{4}\right) + f\left(A + \frac{5}{2} \frac{B-A}{4}\right) \right. \\ & \quad \left. + f\left(A + \frac{7}{2} \frac{B-A}{4}\right) \right\} \cdot \frac{B-A}{4} \\ &= \left\{ f\left(\frac{7}{8}A + \frac{1}{8}B\right) + f\left(\frac{5}{8}A + \frac{3}{8}B\right) + f\left(\frac{3}{8}A + \frac{5}{8}B\right) \right. \\ & \quad \left. + f\left(\frac{1}{8}A + \frac{7}{8}B\right) \right\} \cdot \frac{B-A}{4} \end{aligned}$$

Now recalling that $A = b$ and $B = a$, we have that the midpoint Riemann sum approximation to $\int_b^a f(x) dx$ with 4 subintervals is

$$\left\{ f\left(\frac{7}{8}b + \frac{1}{8}a\right) + f\left(\frac{5}{8}b + \frac{3}{8}a\right) + f\left(\frac{3}{8}b + \frac{5}{8}a\right) + f\left(\frac{1}{8}b + \frac{7}{8}a\right) \right\} \cdot \frac{a-b}{4}$$

Thus we see that the Riemann sums for the two integrals are nearly identical — the only difference being the factor of $\frac{b-a}{4}$ versus $\frac{a-b}{4}$. Hence the two Riemann sums are negatives of each other. The same computation with n subintervals shows that the midpoint Riemann sum approximations to $\int_b^a f(x) dx$ and $\int_a^b f(x) dx$ with n subintervals are negatives of each other. Taking the limit $n \rightarrow \infty$ gives $\int_b^a f(x) dx = -\int_a^b f(x) dx$.

- Finally consider (c) — we will not give a formal proof of this, but instead will interpret it geometrically. Indeed one can also interpret (a) geometrically. In both cases these become statements about areas:

$$\int_a^a f(x) dx = 0 \quad \text{and} \quad \int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$$

are

$$\text{Area}\{ (x, y) \mid a \leq x \leq a, 0 \leq y \leq f(x) \} = 0$$

and

$$\begin{aligned} & \text{Area}\{ (x, y) \mid a \leq x \leq b, 0 \leq y \leq f(x) \} \\ &= \text{Area}\{ (x, y) \mid a \leq x \leq c, 0 \leq y \leq f(x) \} \\ &+ \text{Area}\{ (x, y) \mid c \leq x \leq b, 0 \leq y \leq f(x) \} \end{aligned}$$

respectively. Both of these geometric statements are intuitively obvious. See the figures below.



Note that we have assumed that $a \leq c \leq b$ and that $f(x) \geq 0$. One can remove these restrictions and also make the proof more formal, but it becomes quite tedious and less intuitive.

Remark 1.2.4

For notational simplicity, let's assume that $a \leq c \leq b$ and $f(x) \geq 0$ for all $a \leq x \leq b$. The geometric interpretations of the identities

$$\int_a^a f(x) dx = 0 \quad \text{and} \quad \int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$$

are

$$\text{Area}\{ (x, y) \mid a \leq x \leq a, 0 \leq y \leq f(x) \} = 0$$

and

$$\begin{aligned} & \text{Area}\{ (x, y) \mid a \leq x \leq b, 0 \leq y \leq f(x) \} \\ &= \text{Area}\{ (x, y) \mid a \leq x \leq c, 0 \leq y \leq f(x) \} \\ &+ \text{Area}\{ (x, y) \mid c \leq x \leq b, 0 \leq y \leq f(x) \} \end{aligned}$$

respectively. Both of these geometric statements are intuitively obvious. See the figures below. We won't give a formal proof.



So we concentrate on the formula $\int_b^a f(x) dx = -\int_a^b f(x) dx$. The midpoint Riemann sum approximation to $\int_a^b f(x) dx$ with 4 subintervals (so that each subinterval has width $\frac{b-a}{4}$) is

$$\begin{aligned} & \left\{ f\left(a + \frac{1}{2} \frac{b-a}{4}\right) + f\left(a + \frac{3}{2} \frac{b-a}{4}\right) + f\left(a + \frac{5}{2} \frac{b-a}{4}\right) + f\left(a + \frac{7}{2} \frac{b-a}{4}\right) \right\} \frac{b-a}{4} \\ &= \left\{ f\left(\frac{7}{8}a + \frac{1}{8}b\right) + f\left(\frac{5}{8}a + \frac{3}{8}b\right) + f\left(\frac{3}{8}a + \frac{5}{8}b\right) + f\left(\frac{1}{8}a + \frac{7}{8}b\right) \right\} \frac{b-a}{4} \end{aligned} \quad (*)$$

We're now going to write out the midpoint Riemann sum approximation to $\int_b^a f(x) dx$ with 4 subintervals. Note that b is now the lower limit on the integral and a is now the upper limit on the integral. This is likely to cause confusion when we write out the Riemann sum, so we'll temporarily rename b to A and a to B . The midpoint Riemann sum approximation to $\int_A^B f(x) dx$ with 4 subintervals is

$$\begin{aligned} & \left\{ f\left(A + \frac{1}{2} \frac{B-A}{4}\right) + f\left(A + \frac{3}{2} \frac{B-A}{4}\right) + f\left(A + \frac{5}{2} \frac{B-A}{4}\right) + f\left(A + \frac{7}{2} \frac{B-A}{4}\right) \right\} \frac{B-A}{4} \\ & = \left\{ f\left(\frac{7}{8}A + \frac{1}{8}B\right) + f\left(\frac{5}{8}A + \frac{3}{8}B\right) + f\left(\frac{3}{8}A + \frac{5}{8}B\right) + f\left(\frac{1}{8}A + \frac{7}{8}B\right) \right\} \frac{B-A}{4} \end{aligned}$$

Now recalling that $A = b$ and $B = a$, we have that the midpoint Riemann sum approximation to $\int_b^a f(x) dx$ with 4 subintervals is

$$\left\{ f\left(\frac{7}{8}b + \frac{1}{8}a\right) + f\left(\frac{5}{8}b + \frac{3}{8}a\right) + f\left(\frac{3}{8}b + \frac{5}{8}a\right) + f\left(\frac{1}{8}b + \frac{7}{8}a\right) \right\} \frac{a-b}{4} \quad (\star\star)$$

The curly brackets in (\star) and $(\star\star)$ are equal to each other — the terms are just in the reverse order. The factors multiplying the curly brackets in (\star) and $(\star\star)$, namely $\frac{b-a}{4}$ and $\frac{a-b}{4}$, are negatives of each other, so $(\star\star) = -(\star)$. The same computation with n subintervals shows that the midpoint Riemann sum approximations to $\int_b^a f(x) dx$ and $\int_a^b f(x) dx$ with n subintervals are negatives of each other. Taking the limit $n \rightarrow \infty$ gives $\int_b^a f(x) dx = -\int_a^b f(x) dx$.

✓ Example 1.2.5 Revisiting Example 1.1.14

Back in Example 1.1.14 we saw that when $b > 0$ $\int_0^b x dx = \frac{b^2}{2}$. We'll now verify that $\int_0^b x dx = \frac{b^2}{2}$ is still true when $b = 0$ and also when $b < 0$.

- First consider $b = 0$. Then the statement $\int_0^b x dx = \frac{b^2}{2}$ becomes

$$\int_0^0 x dx = 0$$

This is an immediate consequence of Theorem 1.2.3(a).

- Now consider $b < 0$. Let us write $B = -b$, so that $B > 0$. In Example 1.1.14 we saw that

$$\int_{-B}^0 x dx = -\frac{B^2}{2}.$$

So we have

$$\begin{aligned} \int_0^b x dx &= \int_0^{-B} x dx = -\int_{-B}^0 x dx && \text{by Theorem 1.2.3(b)} \\ &= -\left(-\frac{B^2}{2}\right) && \text{by Example 1.1.14} \\ &= \frac{B^2}{2} = \frac{b^2}{2} \end{aligned}$$

We have now shown that

$$\int_0^b x dx = \frac{b^2}{2} \quad \text{for all real numbers } b$$

✓ Example 1.2.6 $\int_a^b x dx$

Applying Theorem 1.2.3 yet again, we have, for all real numbers a and b ,

$$\begin{aligned} \int_a^b x \, dx &= \int_a^0 x \, dx + \int_0^b x \, dx && \text{by Theorem 1.2.3(c) with } c = 0 \\ &= \int_0^b x \, dx - \int_0^a x \, dx && \text{by Theorem 1.2.3(b)} \\ &= \frac{b^2 - a^2}{2} && \text{by Example 1.2.5, twice} \end{aligned}$$

We can also understand this result geometrically.



- (left) When $0 < a < b$, the integral represents the area in green which is the difference of two right-angle triangles — the larger with area $b^2/2$ and the smaller with area $a^2/2$.
- (center) When $a < 0 < b$, the integral represents the signed area of the two displayed triangles. The one above the axis has area $b^2/2$ while the one below has area $-a^2/2$ (since it is below the axis).
- (right) When $a < b < 0$, the integral represents the signed area in purple of the difference between the two triangles — the larger with area $-a^2/2$ and the smaller with area $-b^2/2$.

Theorem 1.2.3(c) shows us how we can split an integral over a larger interval into one over two (or more) smaller intervals. This is particularly useful for dealing with piece-wise functions, like $|x|$.

✓ Example 1.2.7 Integrals involving $|x|$

Using Theorem 1.2.3, we can readily evaluate integrals involving $|x|$. First, recall that

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

Now consider (for example) $\int_{-2}^3 |x| \, dx$. Since the integrand changes at $x = 0$, it makes sense to split the interval of integration at that point:

$$\begin{aligned} \int_{-2}^3 |x| \, dx &= \int_{-2}^0 |x| \, dx + \int_0^3 |x| \, dx && \text{by Theorem 1.2.3} \\ &= \int_{-2}^0 (-x) \, dx + \int_0^3 x \, dx && \text{by definition of } |x| \\ &= -\int_{-2}^0 x \, dx + \int_0^3 x \, dx && \text{by Theorem 1.2.1(c)} \\ &= -(-2^2/2) + (3^2/2) = (4 + 9)/2 \\ &= 13/2 \end{aligned}$$

We can go further still — given a function $f(x)$ we can rewrite the integral of $f(|x|)$ in terms of the integral of $f(x)$ and $f(-x)$.

$$\begin{aligned} \int_{-1}^1 f(|x|) \, dx &= \int_{-1}^0 f(|x|) \, dx + \int_0^1 f(|x|) \, dx \\ &= \int_{-1}^0 f(-x) \, dx + \int_0^1 f(x) \, dx \end{aligned}$$

Here is a more concrete example.

✓ Example 1.2.8 Revisiting Example 1.1.15

Let us compute $\int_{-1}^1 (1 - |x|) \, dx$ again. In Example 1.1.15 we evaluated this integral by interpreting it as the area of a triangle. This time we are going to use *only* the properties given in Theorems 1.2.1 and 1.2.3 and the facts that

$$\int_a^b dx = b - a \quad \text{and} \quad \int_a^b x \, dx = \frac{b^2 - a^2}{2}$$

That $\int_a^b dx = b - a$ is part (e) of Theorem 1.2.1. We saw that $\int_a^b x dx = \frac{b^2 - a^2}{2}$ in Example 1.2.6.

First we are going to get rid of the absolute value signs by splitting the interval over which we integrate. Recalling that $|x| = x$ whenever $x \geq 0$ and $|x| = -x$ whenever $x \leq 0$, we split the interval by Theorem 1.2.3(c),

$$\begin{aligned} \int_{-1}^1 (1 - |x|) dx &= \int_{-1}^0 (1 - |x|) dx + \int_0^1 (1 - |x|) dx \\ &= \int_{-1}^0 (1 - (-x)) dx + \int_0^1 (1 - x) dx \\ &= \int_{-1}^0 (1 + x) dx + \int_0^1 (1 - x) dx \end{aligned}$$

Now we apply parts (a) and (b) of Theorem 1.2.1, and then

$$\begin{aligned} \int_{-1}^1 [1 - |x|] dx &= \int_{-1}^0 1 dx + \int_{-1}^0 x dx + \int_0^1 1 dx - \int_0^1 x dx \\ &= [0 - (-1)] + \frac{0^2 - (-1)^2}{2} + [1 - 0] - \frac{1^2 - 0^2}{2} \\ &= 1 \end{aligned}$$

More properties of integration: even and odd functions

Recall² the following definition

Definition 1.2.9

Let $f(x)$ be a function. Then,

- we say that $f(x)$ is even when $f(x) = f(-x)$ for all x , and
- we say that $f(x)$ is odd when $f(x) = -f(-x)$ for all x .

Of course most functions are neither even nor odd, but many of the standard functions you know are.

✓ Example 1.2.10 Even functions

- Three examples of even functions are $f(x) = |x|$, $f(x) = \cos x$ and $f(x) = x^2$. In fact, if $f(x)$ is any even power of x , then $f(x)$ is an even function.
- The part of the graph $y = f(x)$ with $x \leq 0$, may be constructed by drawing the part of the graph with $x \geq 0$ (as in the figure on the left below) and then reflecting it in the y -axis (as in the figure on the right below).



- In particular, if $f(x)$ is an even function and $a > 0$, then the two sets

$$\begin{aligned} &\{ (x, y) \mid 0 \leq x \leq a \text{ and } y \text{ is between } 0 \text{ and } f(x) \} \\ &\{ (x, y) \mid -a \leq x \leq 0 \text{ and } y \text{ is between } 0 \text{ and } f(x) \} \end{aligned}$$

are reflections of each other in the y -axis and so have the same signed area. That is

$$\int_0^a f(x) dx = \int_{-a}^0 f(x) dx$$

✓ Example 1.2.11 Odd functions

- Three examples of odd functions are $f(x) = \sin x$, $f(x) = \tan x$ and $f(x) = x^3$. In fact, if $f(x)$ is any odd power of x , then $f(x)$ is an odd function.

- The part of the graph $y = f(x)$ with $x \leq 0$, may be constructed by drawing the part of the graph with $x \geq 0$ (like the solid line in the figure on the left below) and then reflecting it in the y -axis (like the dashed line in the figure on the left below) and then reflecting the result in the x -axis (i.e. flipping it upside down, like in the figure on the right, below).



- In particular, if $f(x)$ is an odd function and $a > 0$, then the signed areas of the two sets

$$\begin{aligned} & \{ (x, y) \mid 0 \leq x \leq a \text{ and } y \text{ is between } 0 \text{ and } f(x) \} \\ & \{ (x, y) \mid -a \leq x \leq 0 \text{ and } y \text{ is between } 0 \text{ and } f(x) \} \end{aligned}$$

are negatives of each other — to get from the first set to the second set, you flip it upside down, in addition to reflecting it in the y -axis. That is

$$\int_0^a f(x) dx = - \int_{-a}^0 f(x) dx$$

We can exploit the symmetries noted in the examples above, namely

$$\begin{aligned} \int_0^a f(x) dx &= \int_{-a}^0 f(x) dx && \text{for } f \text{ even} \\ \int_0^a f(x) dx &= - \int_{-a}^0 f(x) dx && \text{for } f \text{ odd} \end{aligned}$$

together with Theorem 1.2.3

$$\int_{-a}^a f(x) dx = \int_{-a}^0 f(x) dx + \int_0^a f(x) dx$$

in order to simplify the integration of even and odd functions over intervals of the form $[-a, a]$.

Theorem 1.2.12 Even and Odd

Let $a > 0$.

- If $f(x)$ is an even function, then

$$\int_{-a}^a f(x) dx = 2 \int_0^a f(x) dx$$

- If $f(x)$ is an odd function, then

$$\int_{-a}^a f(x) dx = 0$$

Proof

For any function

$$\int_{-a}^a f(x) dx = \int_{-a}^0 f(x) dx + \int_0^a f(x) dx$$

When f is even, the two terms on the right hand side are equal. When f is odd, the two terms on the right hand side are negatives of each other.

Optional — More properties of integration: inequalities for integrals

We are still unable to integrate many functions, however with a little work we can infer bounds on integrals from bounds on their integrands.

Theorem 1.2.13 Inequalities for Integrals

Let $a \leq b$ be real numbers and let the functions $f(x)$ and $g(x)$ be integrable on the interval $a \leq x \leq b$.

1. If $f(x) \geq 0$ for all $a \leq x \leq b$, then

$$\int_a^b f(x) dx \geq 0$$

2. If $f(x) \leq g(x)$ for all $a \leq x \leq b$, then

$$\int_a^b f(x) dx \leq \int_a^b g(x) dx$$

3. If there are constants m and M such that $m \leq f(x) \leq M$ for all $a \leq x \leq b$, then

$$m(b-a) \leq \int_a^b f(x) dx \leq M(b-a)$$

4. We have

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx$$

Proof

1. By interpreting the integral as the signed area, this statement simply says that if the curve $y = f(x)$ lies above the x -axis and $a \leq b$, then the signed area of $\{(x, y) \mid a \leq x \leq b, 0 \leq y \leq f(x)\}$ is at least zero. This is quite clear. Alternatively, we could argue more algebraically from Definition 1.1.9. We observe that when we define $\int_a^b f(x) dx$ via Riemann sums, every summand, $f(x_{i,n}^*) \frac{b-a}{n} \geq 0$. Thus the whole sum is nonnegative and consequently, so is the limit, and thus so is the integral.

2. We are assuming that $g(x) - f(x) \geq 0$, so part (a) gives

$$\begin{aligned} \int_a^b [g(x) - f(x)] dx \geq 0 &\implies \int_a^b g(x) dx - \int_a^b f(x) dx \geq 0 \\ &\implies \int_a^b f(x) dx \leq \int_a^b g(x) dx \end{aligned}$$

3. Applying part (b) with $g(x) = M$ for all $a \leq x \leq b$ gives

$$\int_a^b f(x) dx \leq \int_a^b M dx = M(b-a)$$

Similarly, viewing m as a (constant) function, and applying part (b) gives

$$m \leq f(x) \implies \int_a^b \overbrace{m}^{=m(b-a)} dx \leq \int_a^b f(x) dx$$

4. For any x , $|f(x)|$ is either $f(x)$ or $-f(x)$ (depending on whether $f(x)$ is positive or negative), so we certainly have

$$f(x) \leq |f(x)| \quad \text{and} \quad -f(x) \leq |f(x)|$$

Applying part (c) to each of those inequalities gives

$$\int_a^b f(x) dx \leq \int_a^b |f(x)| dx \quad \text{and} \quad -\int_a^b f(x) dx \leq \int_a^b |f(x)| dx$$

Now $\left| \int_a^b f(x) dx \right|$ is either equal to $\int_a^b f(x) dx$ or $-\int_a^b f(x) dx$ (depending on whether the integral is positive or negative). In either case we can apply the above two inequalities to get the same result, namely

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx.$$

✓ Example 1.2.14 $\int_0^{\frac{\pi}{3}} \sqrt{\cos x} dx$

Consider the integral

$$\int_0^{\frac{\pi}{3}} \sqrt{\cos x} dx$$

This is not so easy to compute exactly³ but we can bound it quite quickly.

For x between 0 and $\frac{\pi}{3}$, the function $\cos x$ takes values⁴ between 1 and $\frac{1}{2}$. Thus the function $\sqrt{\cos x}$ takes values between 1 and $\frac{1}{\sqrt{2}}$. That is

$$\frac{1}{\sqrt{2}} \leq \sqrt{\cos x} \leq 1 \quad \text{for } 0 \leq x \leq \frac{\pi}{3}.$$

Consequently, by Theorem 1.2.13(b) with $a = 0$, $b = \frac{\pi}{3}$, $m = \frac{1}{\sqrt{2}}$ and $M = 1$,

$$\frac{\pi}{3\sqrt{2}} \leq \int_0^{\frac{\pi}{3}} \sqrt{\cos x} dx \leq \frac{\pi}{3}$$

Plugging these expressions into a calculator gives us

$$0.7404804898 \leq \int_0^{\frac{\pi}{3}} \sqrt{\cos x} dx \leq 1.047197551$$

Exercises

Stage 1

? 1

For each of the following properties of definite integrals, draw a picture illustrating the concept, interpreting definite integrals as areas under a curve.

For simplicity, you may assume that $a \leq c \leq b$, and that $f(x), g(x)$ give positive values.

- $\int_a^a f(x) dx = 0$, (Theorem 1.2.3, part (a))
- $\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$, (Theorem 1.2.3, part (c))
- $\int_a^b (f(x) + g(x)) dx = \int_a^b f(x) dx + \int_a^b g(x) dx$, (Theorem 1.2.1, part (a))

? 2

If $\int_0^b \cos x dx = \sin b$, then what is $\int_a^b \cos x dx$?

? 3 (*)

Decide whether each of the following statements is true or false. If false, provide a counterexample. If true, provide a brief justification. (Assume that $f(x)$ and $g(x)$ are continuous functions.)

- $\int_{-3}^{-2} f(x) dx = -\int_3^2 f(x) dx.$
- If $f(x)$ is an odd function, then $\int_{-3}^{-2} f(x) dx = \int_2^3 f(x) dx.$
- $\int_0^1 f(x) \cdot g(x) dx = \int_0^1 f(x) dx \cdot \int_0^1 g(x) dx.$

? 4

Suppose we want to make a right Riemann sum with 100 intervals to approximate $\int_5^0 f(x) dx$, where $f(x)$ is a function that gives only positive values.

- What is Δx ?
- Are the heights of our rectangles positive or negative?
- Is our Riemann sum positive or negative?
- Is the signed area under the curve $y = f(x)$ from $x = 0$ to $x = 5$ positive or negative?

Stage 2

? 5 (*)

Suppose $\int_2^3 f(x) dx = -1$ and $\int_2^3 g(x) dx = 5$. Evaluate $\int_2^3 (6f(x) - 3g(x)) dx.$

? 6 (*)

If $\int_0^2 f(x) dx = 3$ and $\int_0^2 g(x) dx = -4$, calculate $\int_0^2 (2f(x) + 3g(x)) dx.$

? 7 (*)

The functions $f(x)$ and $g(x)$ obey

$$\begin{aligned} \int_0^{-1} f(x) dx &= 1 & \int_0^2 f(x) dx &= 2 \\ \int_{-1}^0 g(x) dx &= 3 & \int_0^2 g(x) dx &= 4 \end{aligned}$$

Find $\int_{-1}^2 [3g(x) - f(x)] dx.$

? 8

In Question 1.1.8.45, Section 1.1, we found that

$$\int_0^a \sqrt{1-x^2} dx = \frac{\pi}{4} - \frac{1}{2} \arccos(a) + \frac{1}{2} a \sqrt{1-a^2}$$

when $0 \leq a \leq 1$.

Using this fact, evaluate the following:

- $\int_a^0 \sqrt{1-x^2} dx$, where $-1 \leq a \leq 0$
- $\int_a^1 \sqrt{1-x^2} dx$, where $0 \leq a \leq 1$

? 9 (*)

Evaluate $\int_{-1}^2 |2x| dx$.

You may use the result from Example 1.2.6 that $\int_a^b x dx = \frac{b^2 - a^2}{2}$.

? 10

Evaluate $\int_{-5}^5 x|x| dx$.

? 11

Suppose $f(x)$ is an even function and $\int_{-2}^2 f(x) dx = 10$. What is $\int_{-2}^0 f(x) dx$?

Stage 3

? 12 (*)

Evaluate $\int_{-2}^2 (5 + \sqrt{4 - x^2}) dx$.

? 13 (*)

Evaluate $\int_{-2012}^{+2012} \frac{\sin x}{\log(3 + x^2)} dx$.

? 14 (*)

Evaluate $\int_{-2012}^{+2012} x^{1/3} \cos x dx$.

? 15

Evaluate $\int_0^6 (x - 3)^3 dx$.

? 16

We want to compute the area of an ellipse, $(ax)^2 + (by)^2 = 1$ for some (let's say positive) constants a and b .

1. Solve the equation for the upper half of the ellipse. It should have the form " $y = \dots$ ".
2. Write an integral for the area of the upper half of the ellipse. Using properties of integrals, make the integrand look like the upper half of a circle.
3. Using geometry and your answer to part (b), find the area of the ellipse.

? 17

Fill in the following table: the product of an (even/odd) function with an (even/odd) function is an (even/odd) function. You may assume that both functions are defined for all real numbers.

×	even	odd
---	------	-----

even		
odd		

? 18

Suppose $f(x)$ is an odd function and $g(x)$ is an even function, both defined at $x = 0$. What are the possible values of $f(0)$ and $g(0)$?

? 19

Suppose $f(x)$ is a function defined on all real numbers that is both even and odd. What could $f(x)$ be?

? 20

Is the derivative of an even function even or odd? Is the derivative of an odd function even or odd?

1. Now is a good time to look back at Theorem 1.1.5.
2. We haven't done this in this course, but you should have seen it in your differential calculus course or perhaps even earlier.
3. It is not too hard to use Riemann sums and a computer to evaluate it numerically: $0.948025319\dots$.
4. You know the graphs of sine and cosine, so you should be able to work this out without too much difficulty.

This page titled [1.2: Basic properties of the definite integral](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.3: The Fundamental Theorem of Calculus

We have spent quite a few pages (and lectures) talking about definite integrals, what they are (Definition 1.1.9), when they exist (Theorem 1.1.10), how to compute some special cases (Section 1.1.5), some ways to manipulate them (Theorem 1.2.1 and 1.2.3) and how to bound them (Theorem 1.2.13). Conspicuously missing from all of this has been a discussion of how to compute them in general. It is high time we rectified that.

The single most important tool used to evaluate integrals is called “the fundamental theorem of calculus”. Its grand name is justified — it links the two branches of calculus by connecting derivatives to integrals. In so doing it also tells us how to compute integrals. Very roughly speaking the derivative of an integral is the original function. This fact allows us to compute integrals using antiderivatives ¹. Of course “very rough” is not enough — let's be precise.

Theorem 1.3.1 Fundamental Theorem of Calculus

Let $a < b$ and let $f(x)$ be a function which is defined and continuous on $[a, b]$.

- **Part 1.** Let $F(x) = \int_a^x f(t) dt$ for any $x \in [a, b]$. Then the function $F(x)$ is differentiable and further

$$F'(x) = f(x)$$

- **Part 2.** Let $G(x)$ be any function which is defined and continuous on $[a, b]$. Further let $G(x)$ be differentiable with $G'(x) = f(x)$ for all $a < x < b$. Then

$$\int_a^b f(x) dx = G(b) - G(a) \quad \text{or equivalently} \quad \int_a^b G'(x) dx = G(b) - G(a)$$

Before we prove this theorem and look at a bunch of examples of its application, it is important that we recall one definition from differential calculus — antiderivatives. If $F'(x) = f(x)$ on some interval, then $F(x)$ is called an antiderivative of $f(x)$ on that interval. So Part 2 of the the fundamental theorem of calculus tells us how to evaluate the definite integral of $f(x)$ in terms of any of its antiderivatives — if $G(x)$ is any antiderivative of $f(x)$ then

$$\int_a^b f(x) dx = G(b) - G(a)$$

The form $\int_a^b G'(x) dx = G(b) - G(a)$ of the fundamental theorem relates the rate of change of $G(x)$ over the interval $a \leq x \leq b$ to the net change of G between $x = a$ and $x = b$. For that reason, it is sometimes called the “net change theorem”.

We'll start with a simple example. Then we'll see why the fundamental theorem is true and then we'll do many more, and more involved, examples.

Example 1.3.2 A first example

Consider the integral $\int_a^b x dx$ which we have explored previously in Example 1.2.6.

- The integrand is $f(x) = x$.
- We can readily verify that $G(x) = \frac{x^2}{2}$ satisfies $G'(x) = f(x)$ and so is an antiderivative of the integrand.
- Part 2 of Theorem 1.3.1 then tells us that

$$\int_a^b f(x) dx = G(b) - G(a)$$

$$\int_a^b x dx = \frac{b^2}{2} - \frac{a^2}{2}$$

which is precisely the result we obtained (with more work) in Example 1.2.6.

We do not give completely rigorous proofs of the two parts of the theorem — that is not really needed for this course. We just give the main ideas of the proofs so that you can understand why the theorem is true.

Part 1

We wish to show that if

$$F(x) = \int_a^x f(t) dt \quad \text{then} \quad F'(x) = f(x)$$

- Assume that F is the above integral and then consider $F'(x)$. By definition

$$F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h}$$

- To understand this limit, we interpret the terms $F(x)$, $F(x+h)$ as signed areas. To simplify this further, let's only consider the case that f is always nonnegative and that $h > 0$. These restrictions are not hard to remove, but the proof ideas are a bit cleaner if we keep them in place. Then we have

$$\begin{aligned} F(x+h) &= \text{the area of the region } \{ (t, y) \mid a \leq t \leq x+h, 0 \leq y \leq f(t) \} \\ F(x) &= \text{the area of the region } \{ (t, y) \mid a \leq t \leq x, 0 \leq y \leq f(t) \} \end{aligned}$$

- Then the numerator

$$F(x+h) - F(x) = \text{the area of } \{ (t, y) \mid x \leq t \leq x+h, 0 \leq y \leq f(t) \}$$

This is just the more darkly shaded region in the figure



- We will be taking the limit $h \rightarrow 0$. So suppose that h is very small. Then, as t runs from x to $x+h$, $f(t)$ runs only over a very narrow range of values², all close to $f(x)$.
- So the darkly shaded region is almost a rectangle of width h and height $f(x)$ and so has an area which is very close to $f(x)h$. Thus $\frac{F(x+h)-F(x)}{h}$ is very close to $f(x)$.
- In the limit $h \rightarrow 0$, $\frac{F(x+h)-F(x)}{h}$ becomes exactly $f(x)$, which is precisely what we want.

We can make the above more rigorous using the Mean Value Theorem³.

Part 2

We want to show that $\int_a^b f(t) dt = G(b) - G(a)$. To do this we exploit the fact that the derivative of a constant is zero.

- Let

$$H(x) = \int_a^x f(t) dt - G(x) + G(a)$$

Then the result we wish to prove is that $H(b) = 0$. We will do this by showing that $H(x) = 0$ for all x between a and b .

- We first show that $H(x)$ is constant by computing its derivative:

$$H'(x) = \frac{d}{dx} \int_a^x f(t) dt - \frac{d}{dx} (G(x)) + \frac{d}{dx} (G(a))$$

Since $G(a)$ is a constant, its derivative is 0 and by assumption the derivative of $G(x)$ is just $f(x)$, so

$$= \frac{d}{dx} \int_a^x f(t) dt - f(x)$$

Now Part 1 of the theorem tells us that this derivative is just $f(x)$, so

$$= f(x) - f(x) = 0$$

Hence H is constant.

- To determine which constant we just compute $H(a)$:

$$\begin{aligned}
 H(a) &= \int_a^a f(t) dt - G(a) + G(a) \\
 &= \int_a^a f(t) dt && \text{by Theorem 1.2.3(a)} \\
 &= 0
 \end{aligned}$$

as required.

The simple example we did above (Example 1.3.2), demonstrates the application of part 2 of the fundamental theorem of calculus. Before we do more examples (and there will be many more over the coming sections) we should do some examples illustrating the use of part 1 of the fundamental theorem of calculus. Then we'll move on to part 2.

✓ Example 1.3.3 $\frac{d}{dx} \int_0^x t dt$

Consider the integral $\int_0^x t dt$. We know how to evaluate this — it is just Example 1.3.2 with $a = 0$, $b = x$. So we have two ways to compute the derivative. We can evaluate the integral and then take the derivative, or we can apply Part 1 of the fundamental theorem. We'll do both, and check that the two answers are the same.

First, Example 1.3.2 gives

$$F(x) = \int_0^x t dt = \frac{x^2}{2}$$

So of course $F'(x) = x$. Second, Part 1 of the fundamental theorem of calculus tells us that the derivative of $F(x)$ is just the integrand. That is, Part 1 of the fundamental theorem of calculus also gives $F'(x) = x$.

In the previous example we were able to evaluate the integral explicitly, so we did not need the fundamental theorem to determine its derivative. Here is an example that really does require the use of the fundamental theorem.

✓ Example 1.3.4 $\frac{d}{dx} \int_0^x e^{-t^2} dt$

We would like to find $\frac{d}{dx} \int_0^x e^{-t^2} dt$. In the previous example, we were able to compute the corresponding derivative in two ways — we could explicitly compute the integral and then differentiate the result, or we could apply part 1 of the fundamental theorem of calculus. In this example we do not know the integral explicitly. Indeed it is not possible to express⁴ the integral $\int_0^x e^{-t^2} dt$ as a finite combination of standard functions such as polynomials, exponentials, trigonometric functions and so on.

Despite this, we can find its derivative by just applying the first part of the fundamental theorem of calculus with $f(t) = e^{-t^2}$ and $a = 0$. That gives

$$\begin{aligned}
 \frac{d}{dx} \int_0^x e^{-t^2} dt &= \frac{d}{dx} \int_0^x f(t) dt \\
 &= f(x) = e^{-x^2}
 \end{aligned}$$

Let us ratchet up the complexity of the previous example — we can make the limits of the integral more complicated functions. So consider the previous example with the upper limit x replaced by x^2 :

✓ Example 1.3.5 $\frac{d}{dx} \int_0^{x^2} e^{-t^2} dt$

Consider the integral $\int_0^{x^2} e^{-t^2} dt$. We would like to compute its derivative with respect to x using part 1 of the fundamental theorem of calculus.

The fundamental theorem tells us how to compute the derivative of functions of the form $\int_a^x f(t) dt$ but the integral at hand is *not* of the specified form because the upper limit we have is x^2 , rather than x , — so more care is required. Thankfully we can deal with this obstacle with only a little extra work. The trick is to define an auxiliary function by simply changing the upper limit to x . That is, define

$$E(x) = \int_0^x e^{-t^2} dt$$

Then the integral we want to work with is

$$E(x^2) = \int_0^{x^2} e^{-t^2} dt$$

The derivative $E'(x)$ can be found via part 1 of the fundamental theorem of calculus (as we did in Example 1.3.4) and is $E'(x) = e^{-x^2}$. We can then use this fact with the chain rule to compute the derivative we need:

$$\begin{aligned} \frac{d}{dx} \int_0^{x^2} e^{-t^2} dt &= \frac{d}{dx} E(x^2) && \text{use the chain rule} \\ &= 2xE'(x^2) \\ &= 2xe^{-x^4} \end{aligned}$$

What if both limits of integration are functions of x ? We can still make this work, but we have to split the integral using Theorem 1.2.3.

✓ Example 1.3.6 $\frac{d}{dx} \int_x^{x^2} e^{-t^2} dt$

Consider the integral

$$\int_x^{x^2} e^{-t^2} dt$$

As was the case in the previous example, we have to do a little pre-processing before we can apply the fundamental theorem.

This time (by design), not only is the upper limit of integration x^2 rather than x , but the lower limit of integration also depends on x — this is different from the integral $\int_a^x f(t) dt$ in the fundamental theorem where the *lower* limit of integration is a constant.

Fortunately we can use the basic properties of integrals (Theorem 1.2.3(b) and (c)) to split $\int_x^{x^2} e^{-t^2} dt$ into pieces whose derivatives we already know.

$$\begin{aligned} \int_x^{x^2} e^{-t^2} dt &= \int_x^0 e^{-t^2} dt + \int_0^{x^2} e^{-t^2} dt && \text{by Theorem 1.2.3(c)} \\ &= -\int_0^x e^{-t^2} dt + \int_0^{x^2} e^{-t^2} dt && \text{by Theorem 1.2.3(b)} \end{aligned}$$

With this pre-processing, both integrals are of the right form. Using what we have learned in the the previous two examples,

$$\begin{aligned} \frac{d}{dx} \int_x^{x^2} e^{-t^2} dt &= \frac{d}{dx} \left(-\int_0^x e^{-t^2} dt + \int_0^{x^2} e^{-t^2} dt \right) \\ &= -\frac{d}{dx} \int_0^x e^{-t^2} dt + \frac{d}{dx} \int_0^{x^2} e^{-t^2} dt \\ &= -e^{-x^2} + 2xe^{-x^4} \end{aligned}$$

Before we start to work with part 2 of the fundamental theorem, we need a little terminology and notation. First some terminology — you may have seen this definition in your differential calculus course.

 Definition 1.3.7 Antiderivatives

Let $f(x)$ and $F(x)$ be functions. If $F'(x) = f(x)$ on an interval, then we say that $F(x)$ is an antiderivative of $f(x)$ on that interval.

As we saw above, an antiderivative of $f(x) = x$ is $F(x) = x^2/2$ — we can easily verify this by differentiation. Notice that $x^2/2 + 3$ is also an antiderivative of x , as is $x^2/2 + C$ for any constant C . This observation gives us the following simple lemma.

Lemma 1.3.8

Let $f(x)$ be a function and let $F(x)$ be an antiderivative of $f(x)$. Then $F(x) + C$ is also an antiderivative for any constant C . Further, every antiderivative of $f(x)$ must be of this form.

Proof

There are two parts to the lemma and we prove each in turn.

- Let $F(x)$ be an antiderivative of $f(x)$ and let C be some constant. Then

$$\begin{aligned} \frac{d}{dx}(F(x) + C) &= \frac{d}{dx}(F(x)) + \frac{d}{dx}(C) \\ &= f(x) + 0 \end{aligned}$$

since the derivative of a constant is zero, and by definition the derivative of $F(x)$ is just $f(x)$. Thus $F(x) + C$ is also an antiderivative of $f(x)$.

- Now let $F(x)$ and $G(x)$ both be antiderivatives of $f(x)$ — we will show that $G(x) = F(x) + C$ for some constant C . To do this let $H(x) = G(x) - F(x)$. Then

$$\begin{aligned} \frac{d}{dx}H(x) &= \frac{d}{dx}(G(x) - F(x)) = \frac{d}{dx}G(x) - \frac{d}{dx}F(x) \\ &= f(x) - f(x) = 0 \end{aligned}$$

Since the derivative of $H(x)$ is zero, $H(x)$ must be a constant function⁵. Thus $H(x) = G(x) - F(x) = C$ for some constant C and the result follows.

Based on the above lemma we have the following definition.

Definition 1.3.9

The “indefinite integral of $f(x)$ ” is denoted by $\int f(x) dx$ and should be regarded as the general antiderivative of $f(x)$. In particular, if $F(x)$ is an antiderivative of $f(x)$ then

$$\int f(x) dx = F(x) + C$$

where the C is an arbitrary constant. In this context, the constant C is also often called a “constant of integration”.

Now we just need a tiny bit more notation.

Definition 1.3.10

The symbol

$$\int f(x) dx \Big|_a^b$$

denotes the change in an antiderivative of $f(x)$ from $x = a$ to $x = b$. More precisely, let $F(x)$ be any antiderivative of $f(x)$. Then

$$\int f(x) dx \Big|_a^b = F(x) \Big|_a^b = F(b) - F(a)$$

Notice that this notation allows us to write part 2 of the fundamental theorem as

$$\begin{aligned}\int_a^b f(x) dx &= \int f(x) dx \Big|_a^b \\ &= F(x) \Big|_a^b = F(b) - F(a)\end{aligned}$$

Some texts also use an equivalent notation using square brackets:

$$\int_a^b f(x) dx = [F(x)]_a^b = F(b) - F(a).$$

You should be familiar with both notations.

We'll soon develop some strategies for computing more complicated integrals. But for now, we'll try a few integrals that are simple enough that we can just guess the answer. Of course, any antiderivative that we can guess we can also check — simply differentiate the guess and verify you get back to the original function:

$$\frac{d}{dx} \int f(x) dx = f(x).$$

We do these examples in some detail to help us become comfortable finding indefinite integrals.

✓ **Example 1.3.11** Compute the definite integral $\int_1^2 x dx$

Compute the definite integral $\int_1^2 x dx$.

Solution

We have already seen, in Example 1.2.6, that $\int_1^2 x dx = \frac{2^2-1^2}{2} = \frac{3}{2}$. We shall now rederive that result using the fundamental theorem of calculus.

- The main difficulty in this approach is finding the indefinite integral (an antiderivative) of x . That is, we need to find a function $F(x)$ whose derivative is x . So think back to all the derivatives you computed last term⁶ and try to remember a function whose derivative was something like x .
- This shouldn't be too hard — we recall that the derivatives of polynomials are polynomials. More precisely, we know that

$$\frac{d}{dx} x^n = nx^{n-1}$$

So if we want to end up with just $x = x^1$, we need to take $n = 2$. However this gives us

$$\frac{d}{dx} x^2 = 2x$$

- This is pretty close to what we want except for the factor of 2. Since this is a constant we can just divide both sides by 2 to obtain:

$$\begin{aligned}\frac{1}{2} \cdot \frac{d}{dx} x^2 &= \frac{1}{2} \cdot 2x && \text{which becomes} \\ \frac{d}{dx} \frac{x^2}{2} &= x\end{aligned}$$

which is exactly what we need. It tells us that $x^2/2$ is an antiderivative of x .

- Once one has an antiderivative, it is easy to compute the indefinite integral

$$\int x dx = \frac{1}{2}x^2 + C$$

as well as the definite integral:

$$\begin{aligned}\int_1^2 x dx &= \frac{1}{2}x^2 \Big|_1^2 && \text{since } x^2/2 \text{ is an antiderivative of } x \\ &= \frac{1}{2}2^2 - \frac{1}{2}1^2 = \frac{3}{2}\end{aligned}$$

While the previous example could be computed using signed areas, the following example would be very difficult to compute without using the fundamental theorem of calculus.

✓ Example 1.3.12 Compute $\int_0^{\frac{\pi}{2}} \sin x \, dx$

Compute $\int_0^{\frac{\pi}{2}} \sin x \, dx$.

Solution

- Once again, the crux of the solution is guessing the antiderivative of $\sin x$ — that is finding a function whose derivative is $\sin x$.
- The standard derivative that comes closest to $\sin x$ is

$$\frac{d}{dx} \cos x = -\sin x$$

which is the derivative we want, multiplied by a factor of -1 .

- Just as we did in the previous example, we multiply this equation by a constant to remove this unwanted factor:

$$\begin{aligned} (-1) \cdot \frac{d}{dx} \cos x &= (-1) \cdot (-\sin x) && \text{giving us} \\ \frac{d}{dx} (-\cos x) &= \sin x \end{aligned}$$

This tells us that $-\cos x$ is an antiderivative of $\sin x$.

- Now it is straightforward to compute the integral:

$$\begin{aligned} \int_0^{\frac{\pi}{2}} \sin x \, dx &= -\cos x \Big|_0^{\frac{\pi}{2}} && \text{since } -\cos x \text{ is an antiderivative of } \sin x \\ &= -\cos \frac{\pi}{2} + \cos 0 \\ &= 0 + 1 = 1 \end{aligned}$$

✓ Example 1.3.13 Compute $\int_1^2 \frac{1}{x} \, dx$

Find $\int_1^2 \frac{1}{x} \, dx$.

Solution

- Once again, the crux of the solution is guessing a function whose derivative is $\frac{1}{x}$. Our standard way to differentiate powers of x , namely

$$\frac{d}{dx} x^n = nx^{n-1},$$

doesn't work in this case — since it would require us to pick $n = 0$ and this would give

$$\frac{d}{dx} x^0 = \frac{d}{dx} 1 = 0.$$

- Fortunately, we also know⁷ that

$$\frac{d}{dx} \log x = \frac{1}{x}$$

which is exactly the derivative we want.

- We're now ready to compute the prescribed integral.

$$\begin{aligned} \int_1^2 \frac{1}{x} \, dx &= \log x \Big|_1^2 && \text{since } \log x \text{ is an antiderivative of } 1/x \\ &= \log 2 - \log 1 && \text{since } \log 1 = 0 \\ &= \log 2 \end{aligned}$$

✓ Example 1.3.14 $\int_{-2}^{-1} \frac{1}{x} dx$

Find $\int_{-2}^{-1} \frac{1}{x} dx$.

Solution

- As we saw in the last example,

$$\frac{d}{dx} \log x = \frac{1}{x}$$

and if we naively use this here, then we will obtain

$$\int_{-2}^{-1} \frac{1}{x} dx = \log(-1) - \log(-2)$$

which makes no sense since the logarithm is only defined for positive numbers⁸.

- We can work around this problem using a slight variation of the logarithm — $\log |x|$.
 - When $x > 0$, we know that $|x| = x$ and so we have

$$\begin{aligned} \log |x| &= \log x && \text{differentiating gives us} \\ \frac{d}{dx} \log |x| &= \frac{d}{dx} \log x = \frac{1}{x}. \end{aligned}$$

- When $x < 0$ we have that $|x| = -x$ and so

$$\begin{aligned} \log |x| &= \log(-x) && \text{differentiating with the chain rule gives} \\ \frac{d}{dx} \log |x| &= \frac{d}{dx} \log(-x) \\ &= \frac{1}{(-x)} \cdot (-1) = \frac{1}{x} \end{aligned}$$

- Indeed, more generally we should write the indefinite integral of $1/x$ as

$$\int \frac{1}{x} dx = \log |x| + C$$

which is valid for all positive and negative x . It is, however, undefined at $x = 0$.

- We're now ready to compute the prescribed integral.

$$\begin{aligned} \int_{-2}^{-1} \frac{1}{x} dx &= \log |x| \Big|_{-2}^{-1} && \text{since } \log |x| \text{ is an antiderivative of } 1/x \\ &= \log |-1| - \log |-2| = \log 1 - \log 2 \\ &= -\log 2 = \log \frac{1}{2}. \end{aligned}$$

This next example raises a nasty issue that requires a little care. We know that the function $1/x$ is not defined at $x = 0$ — so can we integrate over an interval that contains $x = 0$ and still obtain an answer that makes sense? More generally can we integrate a function over an interval on which that function has discontinuities?

✓ Example 1.3.15 $\int_{-1}^1 \frac{1}{x^2} dx$

Find $\int_{-1}^1 \frac{1}{x^2} dx$.

Solution

Beware that this is a particularly nasty example, which illustrates a booby trap hidden in the fundamental theorem of calculus. The booby trap explodes when the theorem is applied sloppily.

- The sloppy solution starts, as our previous examples have, by finding an antiderivative of the integrand. In this case we know that

$$\frac{d}{dx} \frac{1}{x} = -\frac{1}{x^2}$$

which means that $-x^{-1}$ is an antiderivative of x^{-2} .

- This suggests (if we proceed naively) that

$$\begin{aligned} \int_{-1}^1 x^{-2} dx &= -\frac{1}{x} \Big|_{-1}^1 && \text{since } -1/x \text{ is an antiderivative of } 1/x^2 \\ &= -\frac{1}{1} - \left(-\frac{1}{-1}\right) \\ &= -2 \end{aligned}$$

Unfortunately,

- At this point we should really start to be concerned. This answer cannot be correct. Our integrand, being a square, is positive everywhere. So our integral represents the area of a region above the x -axis and must be positive.
- So what has gone wrong? The flaw in the computation is that the fundamental theorem of calculus, which says that

$$\text{if } F'(x) = f(x) \text{ then } \int_a^b f(x) dx = F(b) - F(a),$$

is *only* applicable when $F'(x)$ exists and equals $f(x)$ for *all* x between a and b .

- In this case $F'(x) = \frac{1}{x^2}$ does not exist for $x = 0$. So we cannot apply the fundamental theorem of calculus as we tried to above.

An integral, like $\int_{-1}^1 \frac{1}{x^2} dx$, whose integrand is undefined somewhere in the domain of integration is called improper. We'll give a more thorough treatment of improper integrals later in the text. For now, we'll just say that the correct way to define (and evaluate) improper integrals is as a limit of well-defined approximating integrals. We shall later see that, not only is $\int_{-1}^1 \frac{1}{x^2} dx$ not negative, it is infinite.

Remark 1.3.16

For completeness we'll show how to evaluate this integral by sneaking up on the point of discontinuity in the interval of integration. As noted above, we will give a fuller explanation of such integrals later in the text.

- Rather than evaluating the integral directly, we will approximate the integral using definite integrals on intervals that avoid the discontinuity. In the current example, the original domain of integration is $-1 \leq x \leq 1$. The domains of integration of the approximating integrals exclude from $[-1, 1]$ small intervals around $x = 0$.
- The shaded area in the figure below illustrates a typical approximating integral, whose domain of integration consists of the original domain of integration, $[-1, 1]$, but with the interval $[-t, T]$ excluded.



The full domain of integration is only recovered in the limit $t, T \rightarrow 0$.

- For this example, the correct computation is


$$\begin{aligned} \int_{-1}^1 \frac{1}{x^2} dx &= \lim_{t \rightarrow 0^+} \int_{-1}^{-t} \frac{1}{x^2} dx + \lim_{T \rightarrow 0^+} \int_T^1 \frac{1}{x^2} dx \\ &= \lim_{t \rightarrow 0^+} \left[-\frac{1}{x} \right]_{-1}^{-t} + \lim_{T \rightarrow 0^+} \left[-\frac{1}{x} \right]_T^1 \\ &= \lim_{t \rightarrow 0^+} \left[\left(-\frac{1}{-t}\right) - \left(-\frac{1}{-1}\right) \right] + \lim_{T \rightarrow 0^+} \left[\left(-\frac{1}{1}\right) - \left(-\frac{1}{T}\right) \right] \\ &= \lim_{t \rightarrow 0^+} \frac{1}{t} + \lim_{T \rightarrow 0^+} \frac{1}{T} - 2 \\ &= +\infty \end{aligned}$$

- We can interpret this to mean that the signed area under the curve x^{-2} between $x = -1$ and $x = 1$ is infinite.


The above examples have illustrated how we can use the fundamental theorem of calculus to convert knowledge of derivatives into knowledge of integrals. We are now in a position to easily build a table of integrals. Here is a short table of the most important derivatives that we know.

$F(x)$	1	x^n	$\sin x$	$\cos x$	$\tan x$	e^x	$\log_e x $	$\arcsin x$	$\arctan x$
$f(x) = F'(x)$		nx^{n-1}	$\cos x$	$-\sin x$	$\sec^2 x$	e^x	$\frac{1}{x}$	$\frac{1}{\sqrt{1-x^2}}$	$\frac{1}{1+x^2}$

Of course we know other derivatives, such as those of $\sec x$ and $\cot x$, however the ones listed above are arguably the most important ones. From this table (with a very little massaging) we can write down a short table of indefinite integrals.

 **Theorem 1.3.17 Important indefinite integrals**

$f(x)$	$F(x) = \int f(x) dx$
1	$x + C$
x^n	$\frac{1}{n+1} x^{n+1} + C$ provided that $n \neq -1$
$\frac{1}{x}$	$\log_e x + C$
e^x	$e^x + C$
$\sin x$	$-\cos x + C$
$\cos x$	$\sin x + C$
$\sec^2 x$	$\tan x + C$
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin x + C$
$\frac{1}{1+x^2}$	$\arctan x + C$

 **Example 1.3.18 Using Theorem 1.3.17 to compute some integrals**

Find the following integrals

- $\int_2^7 e^x dx$
- $\int_{-2}^2 \frac{1}{1+x^2} dx$
- $\int_0^3 (2x^3 + 7x - 2) dx$

Solution

We can proceed with each of these as before — find the antiderivative and then apply the fundamental theorem. The third integral is a little more complicated, but we can split it up into monomials using Theorem 1.2.1 and do each separately.

- An antiderivative of e^x is just e^x , so

$$\int_2^7 e^x dx = e^x \Big|_2^7 = e^7 - e^2 = e^2(e^5 - 1).$$
- An antiderivative of $\frac{1}{1+x^2}$ is $\arctan(x)$, so

$$\int_{-2}^2 \frac{1}{1+x^2} dx = \arctan(x) \Big|_{-2}^2 = \arctan(2) - \arctan(-2)$$

We can simplify this a little further by noting that $\arctan(x)$ is an odd function, so $\arctan(-2) = -\arctan(2)$ and thus our integral is

$$= 2 \arctan(2)$$

3. We can proceed by splitting the integral using Theorem 1.2.1(d)

$$\begin{aligned} \int_0^3 (2x^3 + 7x - 2) dx &= \int_0^3 2x^3 dx + \int_0^3 7x dx - \int_0^3 2 dx \\ &= 2 \int_0^3 x^3 dx + 7 \int_0^3 x dx - 2 \int_0^3 dx \end{aligned}$$

and because we know that $x^4/4$, $x^2/2$, x are antiderivatives of x^3 , x , 1 respectively, this becomes

$$\begin{aligned} &= \left[\frac{x^4}{2} \right]_0^3 + \left[\frac{7x^2}{2} \right]_0^3 - [2x]_0^3 \\ &= \frac{81}{2} + \frac{7 \cdot 9}{2} - 6 \\ &= \frac{81 + 63 - 12}{2} = \frac{132}{2} = 66. \end{aligned}$$

We can also just find the antiderivative of the whole polynomial by finding the antiderivatives of each term of the polynomial and then recombining them. This is equivalent to what we have done above, but perhaps a little neater:

$$\begin{aligned} \int_0^3 (2x^3 + 7x - 2) dx &= \left[\frac{x^4}{2} + \frac{7x^2}{2} - 2x \right]_0^3 \\ &= \frac{81}{2} + \frac{7 \cdot 9}{2} - 6 = 66. \end{aligned}$$

Exercises

Stage 1

Questions 11 through 14 are meant to help reinforce key ideas in the Fundamental Theorem of Calculus and its proof.

So far, we have been able to guess many antiderivatives. Often, however, antiderivatives are very difficult to guess. In Questions 16 through 19, we will find some antiderivatives that might appear in a table of integrals. Coming up with the antiderivative might be quite difficult (strategies to do just that will form a large part of this semester), but *verifying* that your antiderivative is correct is as simple as differentiating.

? 1 (*)

Suppose that $f(x)$ is a function and $F(x) = e^{(x^2-3)} + 1$ is an antiderivative of $f(x)$. Evaluate the definite integral

$$\int_1^{\sqrt{5}} f(x) dx.$$

? 2 (*)

For the function $f(x) = x^3 - \sin 2x$, find its antiderivative $F(x)$ that satisfies $F(0) = 1$.

? 3 (*)

Decide whether each of the following statements is true or false. Provide a brief justification.

1. If $f(x)$ is continuous on $[1, \pi]$ and differentiable on $(1, \pi)$, then $\int_1^\pi f'(x) dx = f(\pi) - f(1)$.

2. $\int_{-1}^1 \frac{1}{x^2} dx = 0$.

3. If f is continuous on $[a, b]$ then $\int_a^b x f(x) dx = x \int_a^b f(x) dx$.

? 4

True or false: an antiderivative of $\frac{1}{x^2}$ is $\log(x^2)$ (where by $\log x$ we mean logarithm base e).

? 5

True or false: an antiderivative of $\cos(e^x)$ is $\frac{\sin(e^x)}{e^x}$.

? 6

Suppose $F(x) = \int_7^x \sin(t^2) dt$. What is the instantaneous rate of change of $F(x)$ with respect to x ?

? 7 (*)

Suppose $F(x) = \int_2^x e^{1/t} dt$. What is the slope of the tangent line to $y = F(x)$ when $x = 3$?

? 8

Suppose $F'(x) = f(x)$. Give two different antiderivatives of $f(x)$.

? 9

In Question 1.1.8.45, Section 1.1, we found that

$$\int_0^a \sqrt{1-x^2} dx = \frac{\pi}{4} - \frac{1}{2} \arccos(a) + \frac{1}{2} a \sqrt{1-a^2}.$$

1. Verify that $\frac{d}{da} \left\{ \frac{\pi}{4} - \frac{1}{2} \arccos(a) + \frac{1}{2} a \sqrt{1-a^2} \right\} = \sqrt{1-a^2}$.
2. Find a function $F(x)$ that satisfies $F'(x) = \sqrt{1-x^2}$ and $F(0) = \pi$.

? 10

Evaluate the following integrals using the Fundamental Theorem of Calculus Part 2, or explain why it does not apply.

1. $\int_{-\pi}^{\pi} \cos x dx$.
2. $\int_{-\pi}^{\pi} \sec^2 x dx$.
3. $\int_{-2}^0 \frac{1}{x+1} dx$.

? 11

As in the proof of the Fundamental Theorem of Calculus, let $F(x) = \int_a^x f(t) dt$. In the diagram below, shade the area corresponding to $F(x+h) - F(x)$.



? 12 (*)

Let $F(x) = \int_0^x f(t)dt$, where $f(t)$ is shown in the graph below, and $0 \leq x \leq 4$.

1. Is $F(0)$ positive, negative, or zero?
2. Where is $F(x)$ increasing and where is it decreasing?



? 13

Let $G(x) = \int_x^0 f(t)dt$, where $f(t)$ is shown in the graph below, and $0 \leq x \leq 4$.

1. Is $G(0)$ positive, negative, or zero?
2. Where is $G(x)$ increasing and where is it decreasing?



? 14

Let $F(x) = \int_a^x t dt$. Using the definition of the derivative, find $F'(x)$.

? 15

Give a continuous function $f(x)$ so that $F(x) = \int_0^x f(t)dt$ is a constant.

? 16

Evaluate and simplify $\frac{d}{dx} \{x \log(ax) - x\}$, where a is some constant and $\log(x)$ is the logarithm base e . What *antiderivative* does this tell you?

? 17

Evaluate and simplify $\frac{d}{dx} \{e^x (x^3 - 3x^2 + 6x - 6)\}$. What *antiderivative* does this tell you?

? 18

Evaluate and simplify $\frac{d}{dx} \left\{ \log|x + \sqrt{x^2 + a^2}| \right\}$, where a is some constant. What *antiderivative* does this tell you?

? 19

Evaluate and simplify $\frac{d}{dx} \left\{ \sqrt{x(a+x)} - a \log(\sqrt{x} + \sqrt{a+x}) \right\}$, where a is some constant. What *antiderivative* does this tell you?

Stage 2

? 20 (*)

Evaluate $\int_0^2 (x^3 + \sin x) dx$.

? 21 (*)

Evaluate $\int_1^2 \frac{x^2 + 2}{x^2} dx$.

? 22

Evaluate $\int \frac{1}{1 + 25x^2} dx$.

? 23

Evaluate $\int \frac{1}{\sqrt{2 - x^2}} dx$.

? 24

Evaluate $\int \tan^2 x dx$.

? 25

Evaluate $\int 3 \sin x \cos x dx$.

? 26

Evaluate $\int \cos^2 x dx$.

? 27 (*)

If

$$F(x) = \int_0^x \log(2 + \sin t) dt \quad \text{and} \quad G(y) = \int_y^0 \log(2 + \sin t) dt$$

find $F'(\frac{\pi}{2})$ and $G'(\frac{\pi}{2})$.

? 28 (*)

Let $f(x) = \int_1^x 100(t^2 - 3t + 2)e^{-t^2} dt$. Find the interval(s) on which f is increasing.

? 29 (*)

If $F(x) = \int_0^{\cos x} \frac{1}{t^3 + 6} dt$, find $F'(x)$.

? 30 (*)

Compute $f'(x)$ where $f(x) = \int_0^{1+x^4} e^{t^2} dt$.

? 31 (*)

Evaluate $\frac{d}{dx} \left\{ \int_0^{\sin x} (t^6 + 8) dt \right\}$.

? 32 (*)

Let $F(x) = \int_0^{x^3} e^{-t} \sin\left(\frac{\pi t}{2}\right) dt$. Calculate $F'(1)$.

? 33 (*)

Find $\frac{d}{du} \left\{ \int_{\cos u}^0 \frac{dt}{1+t^3} \right\}$.

? 34 (*)

Find $f(x)$ if $x^2 = 1 + \int_1^x f(t) dt$.

? 35 (*)

If $x \sin(\pi x) = \int_0^x f(t) dt$ where f is a continuous function, find $f(4)$.

? 36 (*)

Consider the function $F(x) = \int_0^{x^2} e^{-t} dt + \int_{-x}^0 e^{-t^2} dt$.

1. Find $F'(x)$.
2. Find the value of x for which $F(x)$ takes its minimum value.

? 37 (*)

If $F(x)$ is defined by $F(x) = \int_{x^4-x^3}^x e^{\sin t} dt$, find $F'(x)$.

? 38 (*)

Evaluate $\frac{d}{dx} \left\{ \int_{x^5}^{-x^2} \cos(e^t) dt \right\}$.

? 39 (*)

Differentiate $\int_x^{e^x} \sqrt{\sin t} dt$ for $0 < x < \log \pi$.

? 40 (*)

Evaluate $\int_1^5 f(x) dx$, where $f(x) = \begin{cases} 3 & \text{if } x \leq 3 \\ x & \text{if } x \geq 3 \end{cases}$.

Stage 3

? 41 (*)

If $f'(1) = 2$ and $f'(2) = 3$, find $\int_1^2 f'(x)f''(x) dx$.

? 42 (*)

A car traveling at 30 m/s applies its brakes at time $t = 0$, its velocity (in m/s) decreasing according to the formula $v(t) = 30 - 10t$. How far does the car go before it stops?

? 43 (*)

Compute $f'(x)$ where $f(x) = \int_0^{2x-x^2} \log(1+e^t) dt$. Does $f(x)$ have an absolute maximum? Explain.

? 44 (*)

Find the minimum value of $\int_0^{x^2-2x} \frac{dt}{1+t^4}$. Express your answer as an integral.

? 45 (*)

Define the function $F(x) = \int_0^{x^2} \sin(\sqrt{t}) dt$ on the interval $0 < x < 4$. On this interval, where does $F(x)$ have a maximum?

? 46 (*)

Evaluate $\lim_{n \rightarrow \infty} \frac{\pi}{n} \sum_{j=1}^n \sin\left(\frac{j\pi}{n}\right)$ by interpreting it as a limit of Riemann sums.

? 47 (*)

Use Riemann sums to evaluate the limit $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \frac{1}{1 + \frac{j}{n}}$.

? 48

Below is the graph of $y = f(t)$, $-5 \leq t \leq 5$. Define $F(x) = \int_0^x f(t) dt$ for any x in $[-5, 5]$. Sketch $F(x)$.

? 49 (*)

Define $f(x) = x^3 \int_0^{x^3+1} e^t dt$.

1. Find a formula for the derivative $f'(x)$. (Your formula may include in integral sign.)
2. Find the equation of the tangent line to the graph of $y = f(x)$ at $x = -1$.

? 50

Two students calculate $\int f(x) dx$ for some function $f(x)$.

- Student A calculates $\int f(x) dx = \tan^2 x + x + C$
- Student B calculates $\int f(x) dx = \sec^2 x + x + C$

- It is a fact that $\frac{d}{dx}\{\tan^2 x\} = f(x) - 1$

Who ended up with the correct answer?

? 51

Let $F(x) = \int_0^x x^3 \sin(t) dt$.

1. Evaluate $F(3)$.
2. What is $F'(x)$?

? 52

Let $f(x)$ be an even function, defined everywhere, and let $F(x)$ be an antiderivative of $f(x)$. Is $F(x)$ even, odd, or not necessarily either one? (You may use your answer from Section 1.2, Question 1.2.3.20.)

1. You learned these near the end of your differential calculus course. Now is a good time to revise — but we'll go over them here since they are so important in what follows.
2. Notice that if f were discontinuous, then this might be false.
3. The MVT tells us that there is a number c between x and $x+h$ so that $F'(c) = \frac{F(x+h)-F(x)}{(x+h)-x} = \frac{F(x+h)-F(x)}{h}$. But since $F'(x) = f(x)$, this tells us that $\frac{F(x+h)-F(x)}{h} = f(c)$ where c is trapped between $x+h$ and x . Now when we take the limit as $h \rightarrow 0$ we have that this number c is squeezed to x and the result follows.
4. The integral $\int_0^x e^{-t^2} dt$ is closely related to the “error function” which is an extremely important function in mathematics. While we cannot express this integral (or the error function) as a *finite* combination of polynomials, exponentials etc, we can express it as an infinite series $\int_0^x e^{-t^2} dt = x - \frac{x^3}{3 \cdot 1} + \frac{x^5}{5 \cdot 2} - \frac{x^7}{7 \cdot 3!} + \frac{x^9}{9 \cdot 4!} + \dots + (-1)^k \frac{x^{2k+1}}{(2k+1) \cdot k!} + \dots$. But more on this in Chapter 3.
5. This follows from the Mean Value Theorem. Say $H(x)$ were not constant, then there would be two numbers $a < b$ so that $H(a) \neq H(b)$. Then the MVT tells us that there is a number c between a and b so that $H'(c) = \frac{H(b)-H(a)}{b-a}$. Since both numerator and denominator are nonzero, we know the derivative at c is nonzero. But this would contradict the assumption that derivative of H is zero. Hence we cannot have $a < b$ with $H(a) \neq H(b)$ and so $H(x)$ must be constant.
6. Of course, this assumes that you did your differential calculus course last term. If you did that course at a different time then please think back to that point in time. If it is long enough ago that you don't quite remember when it was, then you should probably do some revision of derivatives of simple functions before proceeding further.
7. Recall that in most mathematics courses (especially this one) we use $\log x$ without any indicated base to denote the natural logarithm — the logarithm base e . Many widely used computer languages, like Java, C, Python, MATLAB, \dots , use $\log(x)$ to denote the logarithm base e too. But many texts also use $\ln x$ to denote the natural logarithm $\log x = \log_e x = \ln x$. The reader should be comfortable with all three notations for this function. They should also be aware that in different contexts — such as in chemistry or physics — it is common to use $\log x$ to denote the logarithm base 10, while in computer science often $\log x$ denotes the logarithm base 2. Context is key.
8. This is not entirely true — one can extend the definition of the logarithm to negative numbers, but to do so one needs to understand complex numbers which is a topic beyond the scope of this course.

This page titled [1.3: The Fundamental Theorem of Calculus](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.4: Substitution

In the previous section we explored the fundamental theorem of calculus and the link it provides between definite integrals and antiderivatives. Indeed, integrals with simple integrands are usually evaluated via this link. In this section we start to explore methods for integrating more complicated integrals. We have already seen — via Theorem 1.2.1 — that integrals interact very nicely with addition, subtraction and multiplication by constants:

$$\int_a^b (Af(x) + Bg(x)) dx = A \int_a^b f(x) dx + B \int_a^b g(x) dx$$

for A, B constants. By combining this with the list of indefinite integrals in Theorem 1.3.17, we can compute integrals of linear combinations of simple functions. For example

$$\begin{aligned} \int_1^4 (e^x - 2 \sin x + 3x^2) dx &= \int_1^4 e^x dx - 2 \int_1^4 \sin x dx + 3 \int_1^4 x^2 dx \\ &= \left(e^x + (-2) \cdot (-\cos x) + 3 \frac{x^3}{3} \right) \Big|_1^4 \quad \text{and so on} \end{aligned}$$

Of course there are a great many functions that can be approached in this way, however there are some very simple examples that cannot.

$$\int \sin(\pi x) dx \quad \int x e^x dx \quad \int \frac{x}{x^2 - 5x + 6} dx$$

In each case the integrands are not linear combinations of simpler functions; in order to compute them we need to understand how integrals (and antiderivatives) interact with compositions, products and quotients. We reached a very similar point in our differential calculus course where we understood the linearity of the derivative,

$$\frac{d}{dx} (Af(x) + Bg(x)) = A \frac{df}{dx} + B \frac{dg}{dx},$$

but had not yet seen the chain, product and quotient rules¹. While we will develop tools to find the second and third integrals in later sections, we should really start with how to integrate compositions of functions.

It is important to state up front, that in general one cannot write down the integral of the composition of two functions — even if those functions are simple. This is not because the integral does not exist. Rather it is because the integral cannot be written down as a finite combination of the standard functions we know. A very good example of this, which we encountered in Example 1.3.4, is the composition of e^x and $-x^2$. Even though we know

$$\int e^x dx = e^x + C \quad \text{and} \quad \int -x^2 dx = -\frac{1}{3}x^3 + C$$

there is no simple function that is equal to the indefinite integral

$$\int e^{-x^2} dx.$$

even though the indefinite integral exists. In this way integration is very different from differentiation.

With that caveat out of the way, we can introduce the substitution rule. The substitution rule is obtained by antidifferentiating the chain rule. In some sense it is the chain rule in reverse. For completeness, let us restate the chain rule:

Theorem 1.4.1 The chain rule

Let $F(u)$ and $u(x)$ be differentiable functions and form their composition $F(u(x))$. Then

$$\frac{d}{dx} F(u(x)) = F'(u(x)) \cdot u'(x)$$

Equivalently, if $y(x) = F(u(x))$, then

$$\frac{dy}{dx} = \frac{dF}{du} \cdot \frac{du}{dx}.$$

Consider a function $f(u)$, which has antiderivative $F(u)$. Then we know that

$$\int f(u) du = \int F'(u) du = F(u) + C$$

Now take the above equation and substitute into it $u = u(x)$ — i.e. replace the variable u with any (differentiable) function of x to get

$$\int f(u) du \Big|_{u=u(x)} = F(u(x)) + C$$

But now the right-hand side is a function of x , so we can differentiate it with respect to x to get

$$\frac{d}{dx} F(u(x)) = F'(u(x)) \cdot u'(x)$$

This tells us that $F(u(x))$ is an antiderivative of the function $F'(u(x)) \cdot u'(x) = f(u(x))u'(x)$. Thus we know

$$\int f(u(x)) \cdot u'(x) dx = F(u(x)) + C = \int f(u) du \Big|_{u=u(x)}$$

This is the substitution rule for indefinite integrals.

Theorem 1.4.2 The substitution rule — indefinite integral version

For any differentiable function $u(x)$:

$$\int f(u(x))u'(x) dx = \int f(u) du \Big|_{u=u(x)}$$

In order to apply the substitution rule successfully we will have to write the integrand in the form $f(u(x)) \cdot u'(x)$. To do this we need to make a good choice of the function $u(x)$; after that it is not hard to then find $f(u)$ and $u'(x)$. Unfortunately there is no one strategy for choosing $u(x)$. This can make applying the substitution rule more art than science ². Here we suggest two possible strategies for picking $u(x)$:

1. Factor the integrand and choose one of the factors to be $u'(x)$. For this to work, you must be able to easily find the antiderivative of the chosen factor. The antiderivative will be $u(x)$.
2. Look for a factor in the integrand that is a function with an argument that is more complicated than just “ x ”. That factor will play the role of $f(u(x))$. Choose $u(x)$ to be the complicated argument.

Here are two examples which illustrate each of those strategies in turn.

✓ Example 1.4.3 $\int 9 \sin^8(x) \cos(x) dx$

Consider the integral

$$\int 9 \sin^8(x) \cos(x) dx$$

We want to massage this into the form of the integrand in the substitution rule — namely $f(u(x)) \cdot u'(x)$. Our integrand can be written as the product of the two factors

$$\underbrace{9 \sin^8(x)}_{\text{first factor}} \cdot \underbrace{\cos(x)}_{\text{second factor}}$$

and we start by determining (or guessing) which factor plays the role of $u'(x)$. We can choose $u'(x) = 9 \sin^8(x)$ or $u'(x) = \cos(x)$.

- If we choose $u'(x) = 9 \sin^8(x)$, then antidifferentiating this to find $u(x)$ is really not very easy. So it is perhaps better to investigate the other choice before proceeding further with this one.
- If we choose $u'(x) = \cos(x)$, then we know (Theorem 1.3.17) that $u(x) = \sin(x)$. This also works nicely because it makes the other factor simplify quite a bit $9 \sin^8(x) = 9u^8$. This looks like the right way to go.

So we go with the second choice. Set $u'(x) = \cos(x)$, $u(x) = \sin(x)$, then

$$\begin{aligned} \int 9 \sin^8(x) \cos(x) dx &= \int 9u(x)^8 \cdot u'(x) dx \\ &= \int 9u^8 du \Big|_{u=\sin(x)} && \text{by the substitution rule} \end{aligned}$$

We are now left with the problem of antidifferentiating a monomial; this we can do with Theorem 1.3.17.

$$\begin{aligned} &= (u^9 + C) \Big|_{u=\sin(x)} \\ &= \sin^9(x) + C \end{aligned}$$

Note that $9 \sin^8(x) \cos(x)$ is a function of x . So our answer, which is the indefinite integral of $9 \sin^8(x) \cos(x)$, must also be a function of x . This is why we have substituted $u = \sin(x)$ in the last step of our solution — it makes our solution a function of x .

✓ Example 1.4.4 $\int 3x^2 \cos(x^3) dx$

Evaluate the integral

$$\int 3x^2 \cos(x^3) dx$$

Solution

Again we are going to use the substitution rule and helpfully our integrand is a product of two factors

$$\underbrace{3x^2}_{\text{first factor}} \cdot \underbrace{\cos(x^3)}_{\text{second factor}}$$

The second factor, $\cos(x^3)$ is a function, namely \cos , with a complicated argument, namely x^3 . So we try $u(x) = x^3$. Then $u'(x) = 3x^2$, which is the other factor in the integrand. So the integral becomes

$$\begin{aligned} \int 3x^2 \cos(x^3) dx &= \int u'(x) \cos(u(x)) dx && \text{just swap order of factors} \\ &= \int \cos(u(x)) u'(x) dx && \text{by the substitution rule} \\ &= \int \cos(u) du \Big|_{u=x^3} \\ &= (\sin(u) + C) \Big|_{u=x^3} && \text{using Theorem 1.3.17} \\ &= \sin(x^3) + C \end{aligned}$$

One more — we'll use this to show how to use the substitution rule with definite integrals.

✓ Example 1.4.5 $\int_0^1 e^x \sin(e^x) dx$

Compute

$$\int_0^1 e^x \sin(e^x) dx.$$

Solution

Again we use the substitution rule.

- The integrand is again the product of two factors and we can choose $u'(x) = e^x$ or $u'(x) = \sin(e^x)$.
- If we choose $u'(x) = e^x$ then $u(x) = e^x$ and the other factor becomes $\sin(u)$ — this looks promising. Notice that if we applied the other strategy of looking for a complicated argument then we would arrive at the same choice.
- So we try $u'(x) = e^x$ and $u(x) = e^x$. This gives (if we ignore the limits of integration for a moment)

$$\begin{aligned} \int e^x \sin(e^x) dx &= \int \sin(u(x))u'(x) dx && \text{apply the substitution rule} \\ &= \int \sin(u) du \Big|_{u=e^x} \\ &= (-\cos(u) + C) \Big|_{u=e^x} \\ &= -\cos(e^x) + C \end{aligned}$$

- But what happened to the limits of integration? We can incorporate them now. We have just shown that the indefinite integral is $-\cos(e^x)$, so by the fundamental theorem of calculus

$$\begin{aligned} \int_0^1 e^x \sin(e^x) dx &= [-\cos(e^x)]_0^1 \\ &= -\cos(e^1) - (-\cos(e^0)) \\ &= -\cos(e) + \cos(1) \end{aligned}$$

Theorem 1.4.2, the substitution rule for indefinite integrals, tells us that if $F(u)$ is any antiderivative for $f(u)$, then $F(u(x))$ is an antiderivative for $f(u(x))u'(x)$. So the fundamental theorem of calculus gives us

$$\begin{aligned} \int_a^b f(u(x))u'(x) dx &= F(u(x)) \Big|_{x=a}^{x=b} \\ &= F(u(b)) - F(u(a)) \\ &= \int_{u(a)}^{u(b)} f(u) du && \text{since } F(u) \text{ is an antiderivative for } f(u) \end{aligned}$$

and we have just found

Theorem 1.4.6 The substitution rule — definite integral version

For any differentiable function $u(x)$:

$$\int_a^b f(u(x))u'(x) dx = \int_{u(a)}^{u(b)} f(u) du$$

Notice that to get from the integral on the left hand side to the integral on the right hand side you

- substitute $u(x) \rightarrow u$ and $u'(x) dx \rightarrow du$,
- set the lower limit for the u integral to the value of u (namely $u(a)$) that corresponds to the lower limit of the x integral (namely $x = a$), and
- set the upper limit for the u integral to the value of u (namely $u(b)$) that corresponds to the upper limit of the x integral (namely $x = b$).

Also note that we now have two ways to evaluate definite integrals of the form $\int_a^b f(u(x))u'(x) dx$.

- We can find the indefinite integral $\int f(u(x))u'(x) dx$, using Theorem 1.4.2, and then evaluate the result between $x = a$ and $x = b$. This is what was done in Example 1.4.5.
- Or we can apply Theorem 1.4.2. This entails finding the indefinite integral $\int f(u) du$ and evaluating the result between $u = u(a)$ and $u = u(b)$. This is what we will do in the following example.

✓ Example 1.4.7 $\int_0^1 x^2 \sin(x^3 + 1) dx$

Compute

$$\int_0^1 x^2 \sin(x^3 + 1) dx$$

Solution

- In this example the integrand is already neatly factored into two pieces. While we could deploy either of our two strategies, it is perhaps easier in this case to choose $u(x)$ by looking for a complicated argument.
- The second factor of the integrand is $\sin(x^3 + 1)$, which is the function \sin evaluated at $x^3 + 1$. So set $u(x) = x^3 + 1$, giving $u'(x) = 3x^2$ and $f(u) = \sin(u)$
- The first factor of the integrand is x^2 which is not quite $u'(x)$, however we can easily massage the integrand into the required form by multiplying and dividing by 3:

$$x^2 \sin(x^3 + 1) = \frac{1}{3} \cdot 3x^2 \cdot \sin(x^3 + 1).$$

- We want this in the form of the substitution rule, so we do a little massaging:

$$\begin{aligned} \int_0^1 x^2 \sin(x^3 + 1) dx &= \int_0^1 \frac{1}{3} \cdot 3x^2 \cdot \sin(x^3 + 1) dx \\ &= \frac{1}{3} \int_0^1 \sin(x^3 + 1) \cdot 3x^2 dx \\ &\qquad\qquad\qquad \text{by Theorem 1.2.1(c)} \end{aligned}$$

- Now we are ready for the substitution rule:

$$\frac{1}{3} \int_0^1 \sin(x^3 + 1) \cdot 3x^2 dx = \frac{1}{3} \int_0^1 \underbrace{\sin(x^3 + 1)}_{=f(u(x))} \cdot \underbrace{3x^2}_{=u'(x)} dx$$

Now set $u(x) = x^3 + 1$ and $f(u) = \sin(u)$

$$\begin{aligned} &= \frac{1}{3} \int_0^1 f(u(x))u'(x) dx \\ &= \frac{1}{3} \int_{u(0)}^{u(1)} f(u) du && \text{by the substitution rule} \\ &= \frac{1}{3} \int_1^2 \sin(u) du && \text{since } u(0) = 1 \text{ and } u(1) = 2 \\ &= \frac{1}{3} [-\cos(u)]_1^2 \\ &= \frac{1}{3} (-\cos(2) - (-\cos(1))) \\ &= \frac{\cos(1) - \cos(2)}{3}. \end{aligned}$$

There is another, and perhaps easier, way to view the manipulations in the previous example. Once you have chosen $u(x)$ you

- make the substitution $u(x) \rightarrow u$,
- replace $dx \rightarrow \frac{1}{u'(x)} du$.

In so doing, we take the integral

$$\begin{aligned} \int_a^b f(u(x)) \cdot u'(x) dx &= \int_{u(a)}^{u(b)} f(u) \cdot u'(x) \cdot \frac{1}{u'(x)} du \\ &= \int_{u(a)}^{u(b)} f(u) du \end{aligned} \quad \text{exactly the substitution rule}$$

but we do not have to manipulate the integrand so as to make $u'(x)$ explicit. Let us redo the previous example by this approach.

✓ Example 1.4.8 Example 1.4.7 revisited

Compute the integral

$$\int_0^1 x^2 \sin(x^3 + 1) dx$$

Solution

- We have already observed that one factor of the integrand is $\sin(x^3 + 1)$, which is \sin evaluated at $x^3 + 1$. Thus we try setting $u(x) = x^3 + 1$.
- This makes $u'(x) = 3x^2$, and we replace $u(x) = x^3 + 1 \rightarrow u$ and $dx \rightarrow \frac{1}{u'(x)} du = \frac{1}{3x^2} du$:

$$\begin{aligned} \int_0^1 x^2 \sin(x^3 + 1) dx &= \int_{u(0)}^{u(1)} \underbrace{x^2 \sin(x^3 + 1)}_{=\sin(u)} \frac{1}{3x^2} du \\ &= \int_1^2 \sin(u) \frac{x^2}{3x^2} du \\ &= \int_1^2 \frac{1}{3} \sin(u) du \\ &= \frac{1}{3} \int_1^2 \sin(u) du \end{aligned}$$

which is precisely the integral we found in Example 1.4.7.

✓ Example 1.4.9 Some more substitutions

Compute the indefinite integrals

$$\int \sqrt{2x+1} dx \quad \text{and} \quad \int e^{3x-2} dx$$

Solution

- Starting with the first integral, we see that it is not too hard to spot the complicated argument. If we set $u(x) = 2x + 1$ then the integrand is just \sqrt{u} .
- Hence we substitute $2x + 1 \rightarrow u$ and $dx \rightarrow \frac{1}{u'(x)} du = \frac{1}{2} du$:

$$\begin{aligned} \int \sqrt{2x+1} dx &= \int \sqrt{u} \frac{1}{2} du \\ &= \int u^{1/2} \frac{1}{2} du \\ &= \left(\frac{2}{3} u^{3/2} \cdot \frac{1}{2} + C \right) \Big|_{u=2x+1} \\ &= \frac{1}{3} (2x+1)^{3/2} + C \end{aligned}$$

- We can evaluate the second integral in much the same way. Set $u(x) = 3x - 2$ and replace dx by $\frac{1}{u'(x)} du = \frac{1}{3} du$:

$$\begin{aligned}\int e^{3x-2} dx &= \int e^u \frac{1}{3} du \\ &= \left(\frac{1}{3} e^u + C \right) \Big|_{u=3x-2} \\ &= \frac{1}{3} e^{3x-2} + C\end{aligned}$$

This last example illustrates that substitution can be used to easily deal with arguments of the form $ax + b$, i.e. that are linear functions of x , and suggests the following theorem.

Theorem 1.4.10

Let $F(u)$ be an antiderivative of $f(u)$ and let a, b be constants. Then

$$\int f(ax + b) dx = \frac{1}{a} F(ax + b) + C$$

Proof

We can show this using the substitution rule. Let $u(x) = ax + b$ so $u'(x) = a$, then

$$\begin{aligned}\int f(ax + b) dx &= \int f(u) \cdot \frac{1}{u'(x)} du \\ &= \int \frac{1}{a} f(u) du \\ &= \frac{1}{a} \int f(u) du && \text{since } a \text{ is a constant} \\ &= \frac{1}{a} F(u) \Big|_{u=ax+b} + C && \text{since } F(u) \text{ is an antiderivative of } f(u) \\ &= \frac{1}{a} F(ax + b) + C.\end{aligned}$$

Now we can do the following example using the substitution rule or the above theorem:

✓ Example 1.4.11 $\int_0^{\frac{\pi}{2}} \cos(3x) dx$

Compute $\int_0^{\frac{\pi}{2}} \cos(3x) dx$.

- In this example we should set $u = 3x$, and substitute $dx \rightarrow \frac{1}{u'(x)} du = \frac{1}{3} du$. When we do this we also have to convert the limits of the integral: $u(0) = 0$ and $u(\pi/2) = 3\pi/2$. This gives

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \cos(3x) dx &= \int_0^{\frac{3\pi}{2}} \cos(u) \frac{1}{3} du \\ &= \left[\frac{1}{3} \sin(u) \right]_0^{\frac{3\pi}{2}} \\ &= \frac{\sin(3\pi/2) - \sin(0)}{3} \\ &= \frac{-1 - 0}{3} = -\frac{1}{3}.\end{aligned}$$

- We can also do this example more directly using the above theorem. Since $\sin(x)$ is an antiderivative of $\cos(x)$, Theorem 1.4.10 tells us that $\frac{\sin(3x)}{3}$ is an antiderivative of $\cos(3x)$. Hence

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \cos(3x) dx &= \left[\frac{\sin(3x)}{3} \right]_0^{\frac{\pi}{2}} \\ &= \frac{\sin(3\pi/2) - \sin(0)}{3} \\ &= -\frac{1}{3}.\end{aligned}$$

The rest of this section is just more examples of the substitution rule. We recommend that you after reading these that you practice many examples by yourself under exam conditions.

✓ Example 1.4.12 $\int_0^1 x^2 \sin(1 - x^3) dx$

This integral looks a lot like that of Example 1.4.7. It makes sense to try $u(x) = 1 - x^3$ since it is the argument of $\sin(1 - x^3)$. We

- substitute $u = 1 - x^3$ and
- replace dx with $\frac{1}{u'(x)} du = \frac{1}{-3x^2} du$,
- when $x = 0$, we have $u = 1 - 0^3 = 1$ and
- when $x = 1$, we have $u = 1 - 1^3 = 0$.

So

$$\begin{aligned}\int_0^1 x^2 \sin(1 - x^3) \cdot dx &= \int_1^0 x^2 \sin(u) \cdot \frac{1}{-3x^2} du \\ &= \int_1^0 -\frac{1}{3} \sin(u) du.\end{aligned}$$

Note that the lower limit of the u -integral, namely 1, is larger than the upper limit, which is 0. There is absolutely nothing wrong with that. We can simply evaluate the u -integral in the normal way. Since $-\cos(u)$ is an antiderivative of $\sin(u)$:

$$\begin{aligned}&= \left[\frac{\cos(u)}{3} \right]_1^0 \\ &= \frac{\cos(0) - \cos(1)}{3} \\ &= \frac{1 - \cos(1)}{3}.\end{aligned}$$

✓ Example 1.4.13 $\int_0^1 \frac{1}{(2x+1)^3} dx$

Compute $\int_0^1 \frac{1}{(2x+1)^3} dx$.

We could do this one using Theorem 1.4.10, but its not too hard to do without. We can think of the integrand as the function “one over a cube” with the argument $2x + 1$. So it makes sense to substitute $u = 2x + 1$. That is

- set $u = 2x + 1$ and
- replace $dx \rightarrow \frac{1}{u'(x)} du = \frac{1}{2} du$.
- When $x = 0$, we have $u = 2 \times 0 + 1 = 1$ and
- when $x = 1$, we have $u = 2 \times 1 + 1 = 3$.

So

$$\begin{aligned}
 \int_0^1 \frac{1}{(2x+1)^3} dx &= \int_1^3 \frac{1}{u^3} \cdot \frac{1}{2} du \\
 &= \frac{1}{2} \int_1^3 u^{-3} du \\
 &= \frac{1}{2} \left[\frac{u^{-2}}{-2} \right]_1^3 \\
 &= \frac{1}{2} \left(\frac{1}{-2} \cdot \frac{1}{9} - \frac{1}{-2} \cdot \frac{1}{1} \right) \\
 &= \frac{1}{2} \left(\frac{1}{2} - \frac{1}{18} \right) = \frac{1}{2} \cdot \frac{8}{18} \\
 &= \frac{2}{9}
 \end{aligned}$$

✓ Example 1.4.14 $\int_0^1 \frac{x}{1+x^2} dx$

Evaluate $\int_0^1 \frac{x}{1+x^2} dx$.

Solution

- The integrand can be rewritten as $x \cdot \frac{1}{1+x^2}$. This second factor suggests that we should try setting $u = 1 + x^2$ — and so we interpret the second factor as the function “one over” evaluated at argument $1 + x^2$.
- With this choice we
 - set $u = 1 + x^2$,
 - substitute $dx \rightarrow \frac{1}{2x} du$, and
 - translate the limits of integration: when $x = 0$, we have $u = 1 + 0^2 = 1$ and when $x = 1$, we have $u = 1 + 1^2 = 2$.
- The integral then becomes

$$\begin{aligned}
 \int_0^1 \frac{x}{1+x^2} dx &= \int_1^2 \frac{x}{u} \frac{1}{2x} du \\
 &= \int_1^2 \frac{1}{2u} du \\
 &= \frac{1}{2} [\log |u|]_1^2 \\
 &= \frac{\log 2 - \log 1}{2} = \frac{\log 2}{2}.
 \end{aligned}$$

Remember that we are using the notation “log” for the natural logarithm, i.e. the logarithm with base e . You might also see it written as “ $\ln x$ ”, or with the base made explicit as “ $\log_e x$ ”.

✓ Example 1.4.15 $\int x^3 \cos(x^4 + 2) dx$

Compute the integral $\int x^3 \cos(x^4 + 2) dx$.

Solution

- The integrand is the product of \cos evaluated at the argument $x^4 + 2$ times x^3 , which aside from a factor of 4, is the derivative of the argument $x^4 + 2$.
- Hence we set $u = x^4 + 2$ and then substitute $dx \rightarrow \frac{1}{u'(x)} du = \frac{1}{4x^3} du$.
- Before proceeding further, we should note that this is an indefinite integral so we don't have to worry about the limits of integration. However we do need to make sure our answer is a function of x — we cannot leave it as a function of u .
- With this choice of u , the integral then becomes

$$\begin{aligned}
 \int x^3 \cos(x^4 + 2) dx &= \int x^3 \cos(u) \frac{1}{4x^3} du \Big|_{u=x^4+2} \\
 &= \int \frac{1}{4} \cos(u) du \Big|_{u=x^4+2} \\
 &= \left(\frac{1}{4} \sin(u) + C \right) \Big|_{u=x^4+2} \\
 &= \frac{1}{4} \sin(x^4 + 2) + C.
 \end{aligned}$$

The next two examples are more involved and require more careful thinking.

✓ Example 1.4.16 $\int \sqrt{1+x^2} x^3 dx$

Compute $\int \sqrt{1+x^2} x^3 dx$.

- An obvious choice of u is the argument inside the square root. So substitute $u = 1 + x^2$ and $dx \rightarrow \frac{1}{2x} du$.
- When we do this we obtain

$$\begin{aligned}
 \int \sqrt{1+x^2} \cdot x^3 dx &= \int \sqrt{u} \cdot x^3 \cdot \frac{1}{2x} du \\
 &= \int \frac{1}{2} \sqrt{u} \cdot x^2 du
 \end{aligned}$$

Unlike all our previous examples, we have not cancelled out all of the x 's from the integrand. However before we do the integral with respect to u , the integrand must be expressed solely in terms of u — no x 's are allowed. (Look that integrand on the right hand side of Theorem 1.4.2.)

- But all is not lost. We can rewrite the factor x^2 in terms of the variable u . We know that $u = 1 + x^2$, so this means $x^2 = u - 1$. Substituting this into our integral gives

$$\begin{aligned}
 \int \sqrt{1+x^2} \cdot x^3 dx &= \int \frac{1}{2} \sqrt{u} \cdot x^2 du \\
 &= \int \frac{1}{2} \sqrt{u} \cdot (u - 1) du \\
 &= \frac{1}{2} \int (u^{3/2} - u^{1/2}) du \\
 &= \frac{1}{2} \left(\frac{2}{5} u^{5/2} - \frac{2}{3} u^{3/2} \right) \Big|_{u=x^2+1} + C \\
 &= \left(\frac{1}{5} u^{5/2} - \frac{1}{3} u^{3/2} \right) \Big|_{u=x^2+1} + C \\
 &= \frac{1}{5} (x^2 + 1)^{5/2} - \frac{1}{3} (x^2 + 1)^{3/2} + C.
 \end{aligned}$$

Oof!

- Don't forget that you can always check the answer by differentiating:

$$\begin{aligned}
 &\frac{d}{dx} \left(\frac{1}{5} (x^2 + 1)^{5/2} - \frac{1}{3} (x^2 + 1)^{3/2} + C \right) \\
 &= \frac{d}{dx} \left(\frac{1}{5} (x^2 + 1)^{5/2} \right) - \frac{d}{dx} \left(\frac{1}{3} (x^2 + 1)^{3/2} \right) \\
 &= \frac{1}{5} \cdot 2x \cdot \frac{5}{2} \cdot (x^2 + 1)^{3/2} - \frac{1}{3} \cdot 2x \cdot \frac{3}{2} \cdot (x^2 + 1)^{1/2} \\
 &= x(x^2 + 1)^{3/2} - x(x^2 + 1)^{1/2} \\
 &= x[(x^2 + 1) - 1] \cdot \sqrt{x^2 + 1} \\
 &= x^3 \sqrt{x^2 + 1}.
 \end{aligned}$$

which is the original integrand ✓.

✓ Example 1.4.17 $\int \tan x \, dx$

Evaluate the indefinite integral $\int \tan(x) \, dx$.

Solution

- At first glance there is nothing to manipulate here and so very little to go on. However we can rewrite $\tan x$ as $\frac{\sin x}{\cos x}$, making the integral $\int \frac{\sin x}{\cos x} \, dx$. This gives us more to work with.
- Now think of the integrand as being the product $\frac{1}{\cos x} \cdot \sin x$. This suggests that we set $u = \cos x$ and that we interpret the first factor as the function “one over” evaluated at $u = \cos x$.
- Substitute $u = \cos x$ and $dx \rightarrow \frac{1}{-\sin x} \, du$ to give:

$$\begin{aligned} \int \frac{\sin x}{\cos x} \, dx &= \int \frac{\sin x}{u} \frac{1}{-\sin x} \, du \Big|_{u=\cos x} \\ &= \int -\frac{1}{u} \, du \Big|_{u=\cos x} \\ &= -\log |\cos x| + C && \text{and if we want to go further} \\ &= \log \left| \frac{1}{\cos x} \right| + C \\ &= \log |\sec x| + C. \end{aligned}$$

In all of the above substitution examples we expressed the new integration variable, u , as a function, $u(x)$, of the old integration variable x . It is also possible to express the old integration variable, x , as a function, $x(u)$, of the new integration variable u . We shall see examples of this in Section 1.9.

Exercises

Recall that we are using $\log x$ to denote the logarithm of x with base e . In other courses it is often denoted $\ln x$.

Stage 1

? 1

1. True or False: $\int \sin(e^x) \cdot e^x \, dx = \int \sin(u) \, du \Big|_{u=e^x} = -\cos(e^x) + C$

2. True or False: $\int_0^1 \sin(e^x) \cdot e^x \, dx = \int_0^1 \sin(u) \, du = 1 - \cos(1)$

? 2

Is the following reasoning sound? If not, fix it.

Problem: Evaluate $\int (2x + 1)^2 \, dx$.

Work: We use the substitution $u = 2x + 1$. Then:

$$\begin{aligned} \int (2x + 1)^2 \, dx &= \int u^2 \, du \\ &= \frac{1}{3} u^3 + C \\ &= \frac{1}{3} (2x + 1)^3 + C \end{aligned}$$

? 3

Is the following reasoning sound? If not, fix it.

Problem: Evaluate $\int_1^\pi \frac{\cos(\log t)}{t} dt.$

Work: We use the substitution $u = \log t$, *so* $du = \frac{1}{t} dt$. *Then:*

$$\begin{aligned} \int_1^\pi \frac{\cos(\log t)}{t} dt &= \int_1^\pi \cos(u) du \\ &= \sin(\pi) - \sin(1) = -\sin(1). \end{aligned}$$

? 4

Is the following reasoning sound? If not, fix it.

Problem: Evaluate $\int_0^{\pi/4} x \tan(x^2) dx.$

Work: We begin with the substitution $u = x^2$, *so* $du = 2x dx$:

$$\begin{aligned} \int_0^{\pi/4} x \tan(x^2) dx &= \int_0^{\pi/4} \frac{1}{2} \tan(x^2) \cdot 2x dx \\ &= \int_0^{\pi^2/16} \frac{1}{2} \tan u du \\ &= \frac{1}{2} \int_0^{\pi^2/16} \frac{\sin u}{\cos u} du \end{aligned}$$

Now we use the substitution $v = \cos u$, *so* $dv = -\sin u du$:

$$\begin{aligned} &= \frac{1}{2} \int_{\cos 0}^{\cos(\pi^2/16)} -\frac{1}{v} dv \\ &= -\frac{1}{2} \int_1^{\cos(\pi^2/16)} \frac{1}{v} dv \\ &= -\frac{1}{2} [\log |v|]_1^{\cos(\pi^2/16)} \\ &= -\frac{1}{2} (\log(\cos(\pi^2/16)) - \log(1)) \\ &= -\frac{1}{2} \log(\cos(\pi^2/16)) \end{aligned}$$

? 5 (*)

What is the integral that results when the substitution $u = \sin x$ is applied to the integral $\int_0^{\pi/2} f(\sin x) dx$?

? 6

Let f and g be functions that are continuous and differentiable everywhere. Simplify

$$\int f'(g(x))g'(x) dx - f(g(x)).$$

Stage 2

? 7 (*)

Use substitution to evaluate $\int_0^1 xe^{x^2} \cos(e^{x^2}) dx$.

? 8 (*)

Let $f(t)$ be any function for which $\int_1^8 f(t) dt = 1$. Calculate the integral $\int_1^2 x^2 f(x^3) dx$.

? 9 (*)

Evaluate $\int \frac{x^2}{(x^3 + 31)^{101}} dx$.

? 10 (*)

Evaluate $\int_e^{e^4} \frac{dx}{x \log x}$.

? 11 (*)

Evaluate $\int_0^{\pi/2} \frac{\cos x}{1 + \sin x} dx$.

? 12 (*)

Evaluate $\int_0^{\pi/2} \cos x \cdot (1 + \sin^2 x) dx$.

? 13 (*)

Evaluate $\int_1^3 (2x - 1)e^{x^2 - x} dx$.

? 14 (*)

Evaluate $\int \frac{(x^2 - 4)x}{\sqrt{4 - x^2}} dx$.

? 15

Evaluate $\int \frac{e^{\sqrt{\log x}}}{2x \sqrt{\log x}} dx$.

Stage 3

Questions 18 through 22 can be solved by substitution, but it may not be obvious which substitution will work. In general, when evaluating integrals, it is not always immediately clear which methods are appropriate. If this happens to you, don't despair, and definitely don't give up! Just guess a method and try it. Even if it fails, you'll probably learn something that you can use to make a better guess. ⁴This is also pretty decent life advice.

? 16 (*)

Calculate $\int_{-2}^2 x e^{x^2} dx$.

? 17 (*)

Calculate $\lim_{n \rightarrow \infty} \sum_{j=1}^n \frac{j}{n^2} \sin\left(1 + \frac{j^2}{n^2}\right)$.

? 18

Evaluate $\int_0^1 \frac{u^3}{u^2+1} du$.

? 19

Evaluate $\int \tan^3 \theta d\theta$.

? 20

Evaluate $\int \frac{1}{e^x + e^{-x}} dx$

? 21

Evaluate $\int_0^1 (1-2x)\sqrt{1-x^2} dx$

? 22

Evaluate $\int \tan x \cdot \log(\cos x) dx$

? 23 (*)

Evaluate $\lim_{n \rightarrow \infty} \sum_{j=1}^n \frac{j}{n^2} \cos\left(\frac{j^2}{n^2}\right)$.

? 24 (*)

Calculate $\lim_{n \rightarrow \infty} \sum_{j=1}^n \frac{j}{n^2} \sqrt{1 + \frac{j^2}{n^2}}$.

Using Riemann sums, prove that

$$\int_a^b 2f(2x) dx = \int_{2a}^{2b} f(x) dx$$

1. If your memory of these rules is a little hazy then you really should go back and revise them before proceeding. You will definitely need a good grasp of the chain rule for what follows in this section.
2. Thankfully this does become easier with experience and we recommend that the reader read some examples and then practice a LOT.
3. A good way to remember this last step is that we replace $\frac{du}{dx} dx$ by just du — which looks like we cancelled out the dx terms: $\frac{du}{dx} dx = du$. While using “cancel the dx ” is a good mnemonic (memory aid), you should not think of the derivative $\frac{du}{dx}$ as a fraction — you are not dividing du by dx .

This page titled [1.4: Substitution](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.5: Area between curves

Before we continue our exploration of different methods for integrating functions, we have now have sufficient tools to examine some simple applications of definite integrals. One of the motivations for our definition of “integral” was the problem of finding the area between some curve and the x -axis for x running between two specified values. More precisely

$$\int_a^b f(x) dx$$

is equal to the signed area between the the curve $y = f(x)$, the x -axis, and the vertical lines $x = a$ and $x = b$.

We found the area of this region by approximating it by the union of tall thin rectangles, and then found the exact area by taking the limit as the width of the approximating rectangles went to zero. We can use the same strategy to find areas of more complicated regions in the xy -plane.

As a preview of the material to come, let $f(x) > g(x) > 0$ and $a < b$ and suppose that we are interested in the area of the region

$$S_1 = \{ (x, y) \mid a \leq x \leq b, g(x) \leq y \leq f(x) \}$$

that is sketched in the left hand figure below.



We already know that $\int_a^b f(x) dx$ is the area of the region

$$S_2 = \{ (x, y) \mid a \leq x \leq b, 0 \leq y \leq f(x) \}$$

sketched in the middle figure above and that $\int_a^b g(x) dx$ is the area of the region

$$S_3 = \{ (x, y) \mid a \leq x \leq b, 0 \leq y \leq g(x) \}$$

sketched in the right hand figure above. Now the region S_1 of the left hand figure can be constructed by taking the region S_2 of center figure and removing from it the region S_3 of the right hand figure. So the area of S_1 is exactly

$$\int_a^b f(x) dx - \int_a^b g(x) dx = \int_a^b (f(x) - g(x)) dx$$

This computation depended on the assumption that $f(x) > g(x)$ and, in particular, that the curves $y = g(x)$ and $y = f(x)$ did not cross. If they do cross, as in this figure



then we have to be a lot more careful. The idea is to separate the domain of integration depending on where $f(x) - g(x)$ changes sign — i.e. where the curves intersect. We will illustrate this in Example 1.5.5 below.

Let us start with an example that makes the link to Riemann sums and definite integrals quite explicit.

✓ Example: 1.5.1 The area between $y = 4 - x^2$ and $y = x$

Find the area bounded by the curves $y = 4 - x^2$, $y = x$, $x = -1$ and $x = 1$.

Solution

- Before we do any calculus, it is a very good idea to make a sketch of the area in question. The curves $y = x$, $x = -1$ and $x = 1$ are all straight lines, while the curve $y = 4 - x^2$ is a parabola whose apex is at $(0, 4)$ and then curves down (because of the minus sign in $-x^2$) with x -intercepts at $(\pm 2, 0)$. Putting these together gives



Notice that the curves $y = 4 - x^2$ and $y = x$ intersect when $4 - x^2 = x$, namely when $x = \frac{1}{2}(-1 \pm \sqrt{17}) \approx 1.56, -2.56$. Hence the curve $y = 4 - x^2$ lies above the line $y = x$ for all $-1 \leq x \leq 1$.

- We are to find the area of the shaded region. Each point (x, y) in this shaded region has $-1 \leq x \leq 1$ and $x \leq y \leq 4 - x^2$. When we were defining the integral (way back in Definition 1.1.9) we used a and b to denote the smallest and largest allowed values of x ; let's do that here too. Let's also use $B(x)$ to denote the bottom curve (i.e. to denote the smallest

allowed value of y for a given x) and use $T(x)$ to denote the top curve (i.e. to denote the largest allowed value of y for a given x). So in this example

$$a = -1 \quad b = 1 \quad B(x) = x \quad T(x) = 4 - x^2$$

and the shaded region is

$$\{ (x, y) \mid a \leq x \leq b, B(x) \leq y \leq T(x) \}$$

- We use the same strategy as we used when defining the integral in Section 1.1.4:
 - Pick a natural number n (that we will later send to infinity), then
 - subdivide the region into n narrow slices, each of width $\Delta x = \frac{b-a}{n}$.
 - For each $i = 1, 2, \dots, n$, slice number i runs from $x = x_{i-1}$ to $x = x_i$, and we approximate its area by the area of a rectangle. We pick a number x_i^* between x_{i-1} and x_i and approximate the slice by a rectangle whose top is at $y = T(x_i^*)$ and whose bottom is at $y = B(x_i^*)$.
 - Thus the area of slice i is approximately $[T(x_i^*) - B(x_i^*)] \Delta x$ (as shown in the figure below).



- So the Riemann sum approximation of the area is

$$\text{Area} \approx \sum_{i=1}^n [T(x_i^*) - B(x_i^*)] \Delta x$$

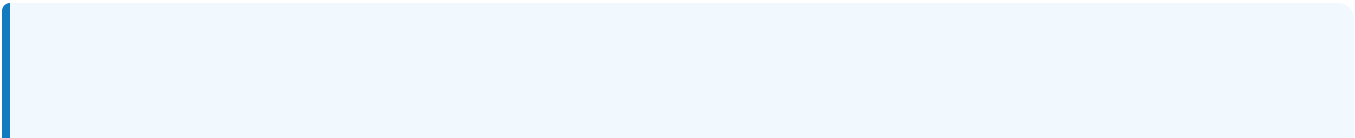
- By taking the limit as $n \rightarrow \infty$ (i.e. taking the limit as the width of the rectangles goes to zero), we convert the Riemann sum into a definite integral (see Definition 1.1.9) and at the same time our approximation of the area becomes the exact area:

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{i=1}^n [T(x_i^*) - B(x_i^*)] \Delta x &= \int_a^b [T(x) - B(x)] dx \\ &\qquad\qquad\qquad \text{Riemann sum} \rightarrow \text{integral} \\ &= \int_{-1}^1 [(4 - x^2) - x] dx \\ &= \int_{-1}^1 [4 - x - x^2] dx \\ &= \left[4x - \frac{x^2}{2} - \frac{x^3}{3} \right]_{-1}^1 \\ &= \left(4 - \frac{1}{2} - \frac{1}{3} \right) - \left(-4 - \frac{1}{2} + \frac{1}{3} \right) \\ &= \frac{24 - 3 - 2}{6} - \frac{-24 - 3 + 2}{6} \\ &= \frac{19}{6} + \frac{25}{6} \\ &= \frac{44}{6} = \frac{22}{3}. \end{aligned}$$

Oof! Thankfully we generally do not need to go through the Riemann sum steps to get to the answer. Usually, provided we are careful to check where curves intersect and which curve lies above which, we can just jump straight to the integral

$$\text{Area} = \int_a^b [T(x) - B(x)] dx.$$

So let us redo the above example.



✓ Example 1.5.2 Example 1.5.1 revisited

Find the area bounded by the curves $y = 4 - x^2$, $y = x$, $x = -1$ and $x = 1$.

Solution

- We first sketch the region



and verify¹ that $y = T(x) = 4 - x^2$ lies above the curve $y = B(x) = x$ on the region $-1 \leq x \leq 1$.

- The area between the curves is then

$$\begin{aligned} \text{Area} &= \int_a^b [T(x) - B(x)] dx \\ &= \int_{-1}^1 [4 - x - x^2] dx \\ &= \left[4x - \frac{x^2}{2} - \frac{x^3}{3} \right]_{-1}^1 \\ &= \frac{19}{6} + \frac{25}{6} = \frac{44}{6} = \frac{22}{3}. \end{aligned}$$

✓ Example 1.5.3 The area between $y = x^2$ and $y = 6x - 2x^2$

Find the area of the finite region bounded by $y = x^2$ and $y = 6x - 2x^2$.

Solution

This is a little different from the previous question, since we are not given bounding lines $x = a$ and $x = b$ — instead we have to determine the minimum and maximum allowed values of x by determining where the curves intersect. Hence our very first task is to get a good idea of what the region looks like by sketching it.

- Start by sketching the region:
 - The curve $y = x^2$ is a parabola. The point on this parabola with the smallest y -coordinate is $(0, 0)$. As $|x|$ increases, y increases so the parabola opens upward.
 - The curve $y = 6x - 2x^2 = -2(x^2 - 3x) = -2(x - \frac{3}{2})^2 + \frac{9}{2}$ is also a parabola. The point on this parabola with the largest value of y has $x = \frac{3}{2}$ (so that the negative term in $-2(x - \frac{3}{2})^2 + \frac{9}{2}$ is zero). So the point with the largest value of y is $(\frac{3}{2}, \frac{9}{2})$. As x moves away from $\frac{3}{2}$, either to the right or to the left, y decreases. So the parabola opens downward. The parabola crosses the x -axis when $0 = 6x - 2x^2 = 2x(3 - x)$. That is, when $x = 0$ and $x = 3$.
 - The two parabolas intersect when $x^2 = 6x - 2x^2$, or

$$\begin{aligned} 3x^2 - 6x &= 0 \\ 3x(x - 2) &= 0 \end{aligned}$$

So there are two points of intersection, one being $x = 0, y = 0^2 = 0$ and the other being $x = 2, y = 2^2 = 4$.

- The finite region between the curves lies between these two points of intersection.

This leads us to the sketch



- So on this region we have $0 \leq x \leq 2$, the top curve is $T(x) = 6x - x^2$ and the bottom curve is $B(x) = x^2$. Hence the area is given by

$$\begin{aligned}
 \text{Area} &= \int_a^b [T(x) - B(x)] dx \\
 &= \int_0^2 [(6x - 2x^2) - (x^2)] dx \\
 &= \int_0^2 [6x - 3x^2] dx \\
 &= \left[6\frac{x^2}{2} - 3\frac{x^3}{3} \right]_0^2 \\
 &= 3(2)^2 - 2^3 = 4
 \end{aligned}$$

✓ Example 1.5.4 The area between $y^2 = 2x + 6$ and $y = x - 1$

Find the area of the finite region bounded by $y^2 = 2x + 6$ and $y = x - 1$.

Solution

We show two different solutions to this problem. The first takes the approach we have in Example 1.5.3 but leads to messy algebra. The second requires a little bit of thinking at the beginning but then is quite straightforward. Before we get to that we should start by sketching the region.

- The curve $y^2 = 2x + 6$, or equivalently $x = \frac{1}{2}y^2 - 3$ is a parabola. The point on this parabola with the smallest x -coordinate has $y = 0$ (so that the positive term in $\frac{1}{2}y^2 - 3$ is zero). So the point on this parabola with the smallest x -coordinate is $(-3, 0)$. As $|y|$ increases, x increases so the parabola opens to the right.
- The curve $y = x - 1$ is a straight line of slope 1 that passes through $x = 1, y = 0$.
- The two curves intersect when $\frac{y^2}{2} - 3 = y + 1$, or

$$\begin{aligned}
 y^2 - 6 &= 2y + 2 \\
 y^2 - 2y - 8 &= 0 \\
 (y + 2)(y - 4) &= 0
 \end{aligned}$$

So there are two points of intersection, one being $y = 4, x = 4 + 1 = 5$ and the other being $y = -2, x = -2 + 1 = -1$.

- Putting this all together gives us the sketch



As noted above, we can find the area of this region by approximating it by a union of narrow vertical rectangles, as we did in Example 1.5.3 — though it is a little harder. The easy way is to approximate it by a union of narrow horizontal rectangles. Just for practice, here is the hard solution. The easy solution is after it.

Harder solution:

- As we have done previously, we approximate the region by a union of narrow vertical rectangles, each of width Δx . Two of those rectangles are illustrated in the sketch



- In this region, x runs from $a = -3$ to $b = 5$. The curve at the top of the region is

$$y = T(x) = \sqrt{2x + 6}$$

The curve at the bottom of the region is more complicated. To the left of $(-1, -2)$ the lower half of the parabola gives the bottom of the region while to the right of $(-1, -2)$ the straight line gives the bottom of the region. So

$$B(x) = \begin{cases} -\sqrt{2x + 6} & \text{if } -3 \leq x \leq -1 \\ x - 1 & \text{if } -1 \leq x \leq 5 \end{cases}$$

- Just as before, the area is still given by the formula $\int_a^b [T(x) - B(x)] dx$, but to accommodate our $B(x)$, we have to split up the domain of integration when we evaluate the integral.

$$\begin{aligned}
 & \int_a^b [T(x) - B(x)] dx \\
 &= \int_{-3}^{-1} [T(x) - B(x)] dx + \int_{-1}^5 [T(x) - B(x)] dx \\
 &= \int_{-3}^{-1} [\sqrt{2x+6} - (-\sqrt{2x+6})] dx + \int_{-1}^5 [\sqrt{2x+6} - (x-1)] dx \\
 &= 2 \int_{-3}^{-1} \sqrt{2x+6} dx + \int_{-1}^5 \sqrt{2x+6} - \int_{-1}^5 (x-1) dx
 \end{aligned}$$

- The third integral is straightforward, while we evaluate the first two via the substitution rule. In particular, set $u = 2x + 6$ and replace $dx \rightarrow \frac{1}{2} du$. Also $u(-3) = 0$, $u(-1) = 4$, $u(5) = 16$. Hence

$$\begin{aligned}
 \text{Area} &= 2 \int_0^4 \sqrt{u} \frac{du}{2} + \int_4^{16} \sqrt{u} \frac{du}{2} - \int_{-1}^5 (x-1) dx \\
 &= 2 \left[\frac{u^{\frac{3}{2}}}{\frac{3}{2}} \right]_0^4 + \left[\frac{u^{\frac{3}{2}}}{\frac{3}{2}} \right]_4^{16} - \left[\frac{x^2}{2} - x \right]_{-1}^5 \\
 &= \frac{2}{3} [8 - 0] + \frac{1}{3} [64 - 8] - \left[\left(\frac{25}{2} - 5 \right) - \left(\frac{1}{2} + 1 \right) \right] \\
 &= \frac{72}{3} - \frac{24}{2} + 6 \\
 &= 18
 \end{aligned}$$

Oof!

Easier solution: The easy way to determine the area of our region is to approximate by narrow horizontal rectangles, rather than narrow vertical rectangles. (Really we are just swapping the roles of x and y in this problem)

- Look at our sketch of the region again — each point (x, y) in our region has $-2 \leq y \leq 4$ and $\frac{1}{2}(y^2 - 6) \leq x \leq y + 1$.
- Let's use
 - c to denote the smallest allowed value of y ,
 - d to denote the largest allowed value of y
 - $L(y)$ (“ L ” stands for “left”) to denote the smallest allowed value of x , when the y -coordinate is y , and
 - $R(y)$ (“ R ” stands for “right”) to denote the largest allowed value of x , when the y -coordinate is y .

So, in this example,

$$c = -2 \quad d = 4 \quad L(y) = \frac{1}{2}(y^2 - 6) \quad R(y) = y + 1$$

and the shaded region is

$$\{ (x, y) \mid c \leq y \leq d, L(y) \leq x \leq R(y) \}$$

- Our strategy is now nearly the same as that used in Example 1.5.1:
 - Pick a natural number n (that we will later send to infinity), then
 - subdivide the interval $c \leq y \leq d$ into n narrow subintervals, each of width $\Delta y = \frac{d-c}{n}$. Each subinterval cuts a thin horizontal slice from the region (see the figure below).
 - We approximate the area of slice number i by the area of a thin horizontal rectangle (indicated by the dark rectangle in the figure below). On this slice, the y -coordinate runs over a very narrow range. We pick a number y_i^* , somewhere in that range. We approximate slice i by a rectangle whose left side is at $x = L(y_i^*)$ and whose right side is at $x = R(y_i^*)$.
 - Thus the area of slice i is approximately $[R(y_i^*) - L(y_i^*)] \Delta y$.



- The desired area is

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \sum_{i=1}^n [R(y_i^*) - L(y_i^*)] \Delta y &= \int_c^d [R(y) - L(y)] dy \\
 &\text{Riemann sum} \rightarrow \text{integral} \\
 &= \int_{-2}^4 [(y+1) - \frac{1}{2}(y^2 - 6)] dy \\
 &= \int_{-2}^4 [-\frac{1}{2}y^2 + y + 4] dy \\
 &= \left[-\frac{1}{6}y^3 + \frac{1}{2}y^2 + 4y \right]_{-2}^4 \\
 &= -\frac{1}{6}(64 - (-8)) + \frac{1}{2}(16 - 4) + 4(4 + 2) \\
 &= -12 + 6 + 24 \\
 &= 18
 \end{aligned}$$

One last example.

✓ Example 1.5.5 Another area

Find the area between the curves $y = \frac{1}{\sqrt{2}}$ and $y = \sin(x)$ with x running from 0 to $\frac{\pi}{2}$.

Solution:

This one is a little trickier since (as we shall see) the region is split into two pieces and we need to treat them separately.

- Again we start by sketching the region.



We want the shaded area.

- Unlike our previous examples, the bounding curves $y = \frac{1}{\sqrt{2}}$ and $y = \sin(x)$ cross in the middle of the region of interest. They cross when $y = \frac{1}{\sqrt{2}}$ and $\sin(x) = y = \frac{1}{\sqrt{2}}$, i.e. when $x = \frac{\pi}{4}$. So
 - to the left of $x = \frac{\pi}{4}$, the top boundary is part of the straight line $y = \frac{1}{\sqrt{2}}$ and the bottom boundary is part of the curve $y = \sin(x)$
 - while to the right of $x = \frac{\pi}{4}$, the top boundary is part of the curve $y = \sin(x)$ and the bottom boundary is part of the straight line $y = \frac{1}{\sqrt{2}}$.
- Thus the formulae for the top and bottom boundaries are

$$\begin{aligned}
 T(x) &= \begin{cases} \frac{1}{\sqrt{2}} & \text{if } 0 \leq x \leq \frac{\pi}{4} \\ \sin(x) & \text{if } \frac{\pi}{4} \leq x \leq \frac{\pi}{2} \end{cases} \\
 B(x) &= \begin{cases} \sin(x) & \text{if } 0 \leq x \leq \frac{\pi}{4} \\ \frac{1}{\sqrt{2}} & \text{if } \frac{\pi}{4} \leq x \leq \frac{\pi}{2} \end{cases}
 \end{aligned}$$

We may compute the area of interest using our canned formula

$$\text{Area} = \int_a^b [T(x) - B(x)] dx$$

but since the formulas for $T(x)$ and $B(x)$ change at the point $x = \frac{\pi}{4}$, we must split the domain of the integral in two at that point².

- Our integral over the domain $0 \leq x \leq \frac{\pi}{2}$ is split into an integral over $0 \leq x \leq \frac{\pi}{4}$ and one over $\frac{\pi}{4} \leq x \leq \frac{\pi}{2}$:

$$\begin{aligned}
 \text{Area} &= \int_0^{\frac{\pi}{2}} [T(x) - B(x)] dx \\
 &= \int_0^{\frac{\pi}{4}} [T(x) - B(x)] dx + \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} [T(x) - B(x)] dx \\
 &= \int_0^{\frac{\pi}{4}} \left[\frac{1}{\sqrt{2}} - \sin(x) \right] dx + \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \left[\sin(x) - \frac{1}{\sqrt{2}} \right] dx \\
 &= \left[\frac{x}{\sqrt{2}} + \cos(x) \right]_0^{\frac{\pi}{4}} + \left[-\cos(x) - \frac{x}{\sqrt{2}} \right]_{\frac{\pi}{4}}^{\frac{\pi}{2}} \\
 &= \left[\frac{1}{\sqrt{2}} \frac{\pi}{4} + \frac{1}{\sqrt{2}} - 1 \right] + \left[\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \frac{\pi}{4} \right] \\
 &= \frac{2}{\sqrt{2}} - 1 \\
 &= \sqrt{2} - 1
 \end{aligned}$$

Exercises

Stage 1

? 1

We want to approximate the area between the graphs of $y = \cos x$ and $y = \sin x$ from $x = 0$ to $x = \pi$ using a left Riemann sum with $n = 4$ rectangles.

1. On the graph below, sketch the four rectangles.
2. Calculate the Riemann approximation.



? 2

We want to approximate the bounded area between the curves $y = \arcsin\left(\frac{2x}{\pi}\right)$ and $y = \sqrt{\frac{\pi x}{2}}$ using $n = 5$ rectangles.

1. Draw the five (vertical) rectangles on the picture below corresponding to a right Riemann sum.
2. Draw five rectangles on the picture below we might use if we were using horizontal rectangles.



? 3 (*)

Write down a definite integral that represents the finite area bounded by the curves $y = x^3 - x$ and $y = x$ for $x \geq 0$. Do not evaluate the integral explicitly.

? 4 (*)

Write down a definite integral that represents the area of the region bounded by the line $y = -\frac{x}{2}$ and the parabola $y^2 = 6 - \frac{5x}{4}$. Do not evaluate the integral explicitly.

? 5 (*)

Write down a definite integral that represents the area of the finite plane region bounded by $y^2 = 4ax$ and $x^2 = 4ay$, where $a > 0$ is a constant. Do not evaluate the integral explicitly.

? 6 (*)

Write down a definite integral that represents the area of the region bounded between the line $x + 12y + 5 = 0$ and the curve $x = 4y^2$. Do not evaluate the integral explicitly.

Stage 2

? 7 (*)

Find the area of the region bounded by the graph of $f(x) = \frac{1}{(2x-4)^2}$ and the x -axis between $x = 0$ and $x = 1$.

? 8 (*)

Find the area between the curves $y = x$ and $y = 3x - x^2$, by first identifying the points of intersection and then integrating.

? 9 (*)

Calculate the area of the region enclosed by $y = 2^x$ and $y = \sqrt{x} + 1$.

? 10 (*)

Find the area of the finite region bounded between the two curves $y = \sqrt{2} \cos(\pi x/4)$ and $y = |x|$.

? 11 (*)

Find the area of the finite region that is bounded by the graphs of $f(x) = x^2 \sqrt{x^3 + 1}$ and $g(x) = 3x^2$.

? 12

Find the area to the left of the y -axis and to the right of the curve $x = y^2 + y$.

? 13 (*)

Find the area of the finite region below $y = \sqrt{9 - x^2}$ and above both $y = |x|$ and $y = \sqrt{1 - x^2}$.

Stage 3

? 14 (*)

The graph below shows the region between $y = 4 + \pi \sin x$ and $y = 4 + 2\pi - 2x$.



Find the area of this region.?

? 15 (*)

Compute the area of the finite region bounded by the curves $x = 0$, $x = 3$, $y = x + 2$ and $y = x^2$.

? 16 (*)

Find the total area between the curves $y = x\sqrt{25 - x^2}$ and $y = 3x$, on the interval $0 \leq x \leq 4$.

? 17

Find the area of the finite region below $y = \sqrt{9 - x^2}$ and $y = x$, and above $y = \sqrt{1 - (x - 1)^2}$.

? 18

Find the area of the finite region bounded by the curve $y = x(x^2 - 4)$ and the line $y = x - 2$.

1. We should do this by checking where the curves intersect; that is by solving $T(x) = B(x)$ and seeing if any of the solutions lie in the range $-1 \leq x \leq 1$.
2. We are effectively computing the area of the region by computing the area of the two disjoint pieces separately. Alternatively, if we set $f(x) = \sin(x)$ and $g(x) = \frac{1}{\sqrt{2}}$, we can rewrite the integral $\int_a^b [T(x) - B(x)] dx$ as $\int_a^b |f(x) - g(x)| dx$. To see that the two integrals are the same, split the domain of integration where $f(x) - g(x)$ changes sign.

This page titled [1.5: Area between curves](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.6: Volumes

Another simple ¹ application of integration is computing volumes. We use the same strategy as we used to express areas of regions in two dimensions as integrals — approximate the region by a union of small, simple pieces whose volume we can compute and then take the limit as the “piece size” tends to zero.

In many cases this will lead to “multivariable integrals” that are beyond our present scope ². But there are some special cases in which this leads to integrals that we can handle. Here are some examples.

✓ Example 1.6.1 Cone

Find the volume of the circular cone of height h and radius r .

Solution

Here is a sketch of the cone.



We have called the vertical axis x , just so that we end up with a “ dx ” integral.

- In what follows we will slice the cone into thin horizontal “pancakes”. In order to approximate the volume of those slices, we need to know the radius of the cone at a height x above its point. Consider the cross sections shown in the following figure.



At full height h , the cone has radius r . If we cut the cone at height x , then by similar triangles (see the figure on the right) the radius will be $\frac{x}{h} \cdot r$.

- Now think of cutting the cone into n thin horizontal “pancakes”. Each such pancake is approximately a squat cylinder of height $\Delta x = \frac{h}{n}$. This is very similar to how we approximated the area under a curve by n tall thin rectangles. Just as we approximated the area under the curve by summing these rectangles, we can approximate the volume of the cone by summing the volumes of these cylinders. Here is a side view of the cone and one of the cylinders.
- We follow the method we used in Example 1.5.1, except that our slices are now pancakes instead of rectangles.
 - Pick a natural number n (that we will later send to infinity), then
 - subdivide the cone into n thin pancakes, each of width $\Delta x = \frac{h}{n}$.
 - For each $i = 1, 2, \dots, n$, pancake number i runs from $x = x_{i-1} = (i-1) \cdot \Delta x$ to $x = x_i = i \cdot \Delta x$, and we approximate its volume by the volume of a squat cone. We pick a number x_i^* between x_{i-1} and x_i and approximate the pancake by a cylinder of height Δx and radius $\frac{x_i^*}{h} r$.
 - Thus the volume of pancake i is approximately $\pi \left(\frac{x_i^*}{h} r \right)^2 \Delta x$ (as shown in the figure above).
- So the Riemann sum approximation of the volume is

$$\text{Area} \approx \sum_{i=1}^n \pi \left(\frac{x_i^*}{h} r \right)^2 \Delta x$$

- By taking the limit as $n \rightarrow \infty$ (i.e. taking the limit as the thickness of the pancakes goes to zero), we convert the Riemann sum into a definite integral (see Definition 1.1.9) and at the same time our approximation of the volume becomes the exact volume:

$$\int_0^h \pi \left(\frac{x}{h} r \right)^2 dx$$

Our life ³ would be easier if we could avoid all this formal work with Riemann sums every time we encounter a new volume. So before we compute the above integral, let us redo the above calculation in a less formal manner.

- Start again from the picture of the cone



and think of slicing it into thin pancakes, each of width dx .

- The pancake at height x above the point of the cone (which is the fraction $\frac{x}{h}$ of the total height of the cone) has
As x runs from 0 to h , the total volume is

$$\begin{aligned} \int_0^h \pi \left(\frac{x}{h}r\right)^2 dx &= \frac{\pi r^2}{h^2} \int_0^h x^2 dx \\ &= \frac{\pi r^2}{h^2} \left[\frac{x^3}{3}\right]_0^h \\ &= \frac{1}{3}\pi r^2 h \end{aligned}$$

- radius $\frac{x}{h} \cdot r$ (the fraction $\frac{x}{h}$ of the full radius, r) and so
- cross-sectional area $\pi \left(\frac{x}{h}r\right)^2$,
- thickness dx — we have done something a little sneaky here, see the discussion below.
- volume $\pi \left(\frac{x}{h}r\right)^2 dx$

In this second computation we are using a time-saving trick. As we saw in the formal computation above, what we really need to do is pick a natural number n , slice the cone into n pancakes each of thickness $\Delta x = \frac{h}{n}$ and then take the limit as $n \rightarrow \infty$. This led to the Riemann sum

$$\sum_{i=1}^n \pi \left(\frac{x_i^*}{h}r\right)^2 \Delta x \quad \text{which becomes} \quad \int_0^h \pi \left(\frac{x}{h}r\right)^2 dx$$

So knowing that we will replace

$$\begin{aligned} \sum_{i=1}^n &\rightarrow \int_0^h \\ x_i^* &\rightarrow x \\ \Delta x &\rightarrow dx \end{aligned}$$

when we take the limit, we have just skipped the intermediate steps. While this is not entirely rigorous, it can be made so, and does save us a lot of algebra.

✓ Example 1.6.2 Sphere

Find the volume of the sphere of radius r .

Solution

We'll find the volume of the part of the sphere in the first octant⁴, sketched below. Then we'll multiply by 8.

- To compute the volume,



we slice it up into thin vertical “pancakes” (just as we did in the previous example).

- Each pancake is one quarter of a thin circular disk. The pancake a distance x from the yz -plane is shown in the sketch above. The radius of that pancake is the distance from the dot shown in the figure to the x -axis, i.e. the y -coordinate of the dot. To get the coordinates of the dot, observe that
 - it lies the xy -plane, and so has z -coordinate zero, and that
 - it also lies on the sphere, so that its coordinates obey $x^2 + y^2 + z^2 = r^2$. Since $z = 0$ and $y > 0$, $y = \sqrt{r^2 - x^2}$.
- So the pancake at distance x from the yz -plane has
 - thickness⁵ dx and
 - radius $\sqrt{r^2 - x^2}$
 - cross-sectional area $\frac{1}{4}\pi(\sqrt{r^2 - x^2})^2$ and hence
 - volume $\frac{\pi}{4}(r^2 - x^2) dx$

- As x runs from 0 to r , the total volume of the part of the sphere in the first octant is

$$\int_0^r \frac{\pi}{4} (r^2 - x^2) dx = \frac{\pi}{4} \left[r^2 x - \frac{x^3}{3} \right]_0^r = \frac{1}{6} \pi r^3$$

and the total volume of the whole sphere is eight times that, which is $\frac{4}{3} \pi r^3$, as expected.

✓ Example 1.6.3 Revolving a region

The region between the lines $y = 3$, $y = 5$, $x = 0$ and $x = 4$ is rotated around the line $y = 2$. Find the volume of the region swept out.

Solution

As with most of these problems, we should start by sketching the problem.



- Consider the region and slice it into thin vertical strips of width dx .
- Now we are to rotate this region about the line $y = 2$. Imagine looking straight down the axis of rotation, $y = 2$, end on. The symbol in the figure above just to the right of the end the line $y = 2$ is supposed to represent your eye⁶. Here is what you see as the rotation takes place.



- Upon rotation about the line $y = 2$ our strip sweeps out a “washer”
 - whose cross-section is a disk of radius $5 - 2 = 3$ from which a disk of radius $3 - 2 = 1$ has been removed so that it has a
 - cross-sectional area of $\pi 3^2 - \pi 1^2 = 8\pi$ and a
 - thickness dx and hence a
 - volume $8\pi dx$.
- As our leftmost strip is at $x = 0$ and our rightmost strip is at $x = 4$, the total

$$\text{Volume} = \int_0^4 8\pi dx = (8\pi)(4) = 32\pi$$

Notice that we could also reach this answer by writing the volume as the difference of two cylinders.

- The outer cylinder has radius $(5 - 2)$ and length 4. This has volume

$$V_{outer} = \pi r^2 \ell = \pi \cdot 3^2 \cdot 4 = 36\pi.$$

- The inner cylinder has radius $(3 - 2)$ and length 4. This has volume

$$V_{inner} = \pi r^2 \ell = \pi \cdot 1^2 \cdot 4 = 4\pi.$$

- The volume we want is the difference of these two, namely

$$V = V_{outer} - V_{inner} = 32\pi.$$

Let us turn up the difficulty a little on this last example.

✓ Example 1.6.4 Revolving again

The region between the curve $y = \sqrt{x}$, and the lines $y = 0$, $x = 0$ and $x = 4$ is rotated around the line $y = 0$. Find the volume of the region swept out.

Solution

We can approach this in much the same way as the previous example.

- Consider the region and cut it into thin vertical strips of width dx .
- When we rotate the region about the line $y = 0$, each strip sweeps out a thin pancake

- whose cross-section is a disk of radius \sqrt{x} with a
- cross-sectional area of $\pi(\sqrt{x})^2 = \pi x$ and a
- thickness dx and hence a
- volume $\pi x dx$.
- As our leftmost strip is at $x = 0$ and our rightmost strip is at $x = 4$, the total


$$\text{Volume} = \int_0^4 \pi x dx = \left[\frac{\pi}{2} x^2 \right]_0^4 = 8\pi$$

In the last example we considered rotating a region around the x -axis. Let us do the same but rotating around the y -axis.

✓ Example 1.6.5 Revolving yet again

The region between the curve $y = \sqrt{x}$, and the lines $y = 0$, $x = 0$ and $x = 4$ is rotated around the line $x = 0$. Find the volume of the region swept out.

Solution:

- We will cut the region into horizontal slices, so we should write x as a function of y . That is, the region is bounded by $x = y^2$, $x = 4$, $y = 0$ and $y = 2$.
- Now slice the region into thin horizontal strips of width dy .

- When we rotate the region about the line $x = 0$, each strip sweeps out a thin washer
 - whose inner radius is y^2 and outer radius is 4, and
 - thickness is dy and hence
 - has volume $\pi(r_{out}^2 - r_{in}^2) dy = \pi(16 - y^4) dy$.
- As our bottommost strip is at $y = 0$ and our topmost strip is at $y = 2$, the total

$$\text{Volume} = \int_0^2 \pi(16 - y^4) dy = \left[16\pi y - \frac{\pi}{5} y^5 \right]_0^2 = 32\pi - \frac{32\pi}{5} = \frac{128\pi}{5}.$$

There is another way⁷ to do this one which we show at the end of this section.

✓ Example 1.6.6 Pyramid

Find the volume of the pyramid which has height h and whose base is a square of side b .

Solution

Here is a sketch of the part of the pyramid that is in the first octant; we display only this portion to make the diagrams simpler.



Note that this diagram shows only 1 quarter of the whole pyramid.

- To compute its volume, we slice it up into thin horizontal “square pancakes”. A typical pancake also appears in the sketch above.
 - The pancake at height z is the fraction $\frac{h-z}{h}$ of the distance from the peak of the pyramid to its base.
 - So the *full* pancake⁸ at height z is a square of side $\frac{h-z}{h}b$. As a check, note that when $z = h$ the pancake has side $\frac{h-h}{h}b = 0$, and when $z = 0$ the pancake has side $\frac{h-0}{h}b = b$.
 - So the pancake has cross-sectional area $\left(\frac{h-z}{h}b\right)^2$ and thickness⁹ dz and hence
 - volume $\left(\frac{h-z}{h}b\right)^2 dz$.
- The volume of the whole pyramid (not just the part of the pyramid in the first octant) is

$$\int_0^h \left(\frac{h-z}{h}b\right)^2 dz = \frac{b^2}{h^2} \int_0^h (h-z)^2 dz$$

Now use the substitution rule with $t = (h - z)$, $dz \rightarrow -dt$

$$\begin{aligned}
 &= \frac{b^2}{h^2} \int_h^0 -t^2 dt \\
 &= -\frac{b^2}{h^2} \left[\frac{t^3}{3} \right]_h^0 \\
 &= -\frac{b^2}{h^2} \left[-\frac{h^3}{3} \right] \\
 &= \frac{1}{3} b^2 h
 \end{aligned}$$

Let's ramp up the difficulty a little.

✓ Example 1.6.7 Napkin Ring

Suppose you make two napkin rings¹⁰ by drilling holes with different diameters through two wooden balls. One ball has radius r and the other radius R with $r < R$. You choose the diameter of the holes so that both napkin rings have the same height, $2h$. See the figure below.



Which¹¹ ring has more wood in it?

Solution

We'll compute the volume of the napkin ring with radius R . We can then obtain the volume of the napkin ring of radius r , by just replacing $R \mapsto r$ in the result.

- To compute the volume of the napkin ring of radius R , we slice it up into thin horizontal “pancakes”. Here is a sketch of the part of the napkin ring in the first octant showing a typical pancake.



- The coordinates of the two points marked in the yz -plane of that figure are found by remembering that
 - the equation of the sphere is $x^2 + y^2 + z^2 = R^2$.
 - The two points have $y > 0$ and are in the yz -plane, so that $x = 0$ for them. So $y = \sqrt{R^2 - z^2}$.
 - In particular, at the top of the napkin ring $z = h$, so that $y = \sqrt{R^2 - h^2}$.
- The pancake at height z , shown in the sketch, is a “washer” — a circular disk with a circular hole cut in its center.
 - The outer radius of the washer is $\sqrt{R^2 - z^2}$ and
 - the inner radius of the washer is $\sqrt{R^2 - h^2}$. So the
 - cross-sectional area of the washer is

$$\pi(\sqrt{R^2 - z^2})^2 - \pi(\sqrt{R^2 - h^2})^2 = \pi(h^2 - z^2)$$

- The pancake at height z
 - has thickness dz and
 - cross-sectional area $\pi(h^2 - z^2)$ and hence
 - volume $\pi(h^2 - z^2) dz$.
- Since z runs from $-h$ to $+h$, the total volume of wood in the napkin ring of radius R is

$$\begin{aligned}
 \int_{-h}^h \pi(h^2 - z^2) dz &= \pi \left[h^2 z - \frac{z^3}{3} \right]_{-h}^h \\
 &= \pi \left[\left(h^3 - \frac{h^3}{3} \right) - \left((-h)^3 - \frac{(-h)^3}{3} \right) \right] \\
 &= \pi \left[\frac{2}{3} h^3 - \frac{2}{3} (-h)^3 \right] \\
 &= \frac{4\pi}{3} h^3
 \end{aligned}$$

This volume is independent of R . Hence the napkin ring of radius r contains precisely the same volume of wood as the napkin ring of radius R !

✓ Example 1.6.8 Notch

A 45° notch is cut to the center of a cylindrical log having radius 20cm. One plane face of the notch is perpendicular to the axis of the log. See the sketch below. What volume of wood was removed?



Solution

We show two solutions to this problem which are of comparable difficulty. The difference lies in the shape of the pancakes we use to slice up the volume. In solution 1 we cut rectangular pancakes parallel to the yz -plane and in solution 2 we slice triangular pancakes parallel to the xz -plane.

Solution 1:

- Concentrate on the notch. Rotate it around so that the plane face lies in the xy -plane.
- Then slice the notch into vertical rectangles (parallel to the yz -plane) as in the figure on the left below.



- The cylindrical log had radius 20cm. So the circular part of the boundary of the base of the notch has equation $x^2 + y^2 = 20^2$. (We're putting the origin of the xy -plane at the center of the circle.) If our coordinate system is such that x is constant on each slice, then
 - the base of the slice is the line segment from $(x, -y, 0)$ to $(x, +y, 0)$ where $y = \sqrt{20^2 - x^2}$ so that
 - the slice has width $2y = 2\sqrt{20^2 - x^2}$ and
 - height x (since the upper face of the notch is at 45° to the base — see the side view sketched in the figure on the right above).
 - So the slice has cross-sectional area $2x\sqrt{20^2 - x^2}$.
- On the base of the notch x runs from 0 to 20 so the volume of the notch is

$$V = \int_0^{20} 2x\sqrt{20^2 - x^2} dx$$

Make the change of variables $u = 20^2 - x^2$ (don't forget to change $dx \rightarrow -\frac{1}{2x} du$):

$$\begin{aligned} V &= \int_{20^2}^0 -\sqrt{u} du \\ &= \left[-\frac{u^{3/2}}{3/2} \right]_{20^2}^0 \\ &= \frac{2}{3} 20^3 = \frac{16,000}{3} \end{aligned}$$

Solution 2:

- Concentrate of the notch. Rotate it around so that its base lies in the xy -plane with the skinny edge along the y -axis.
- Slice the notch into triangles parallel to the xz -plane as in the figure on the left below. In the figure below, the triangle happens to lie in a plane where y is negative.



- The cylindrical log had radius 20cm. So the circular part of the boundary of the base of the notch has equation $x^2 + y^2 = 20^2$. Our coordinate system is such that y is constant on each slice, so that
 - the base of the triangle is the line segment from $(0, y, 0)$ to $(x, y, 0)$ where $x = \sqrt{20^2 - y^2}$ so that
 - the triangle has base $x = \sqrt{20^2 - y^2}$ and
 - height $x = \sqrt{20^2 - y^2}$ (since the upper face of the notch is at 45° to the base — see the side view sketched in the figure on the right above).
 - So the slice has cross-sectional area $\frac{1}{2}(\sqrt{20^2 - y^2})^2$.

- On the base of the notch y runs from -20 to 20 , so the volume of the notch is

$$\begin{aligned} V &= \frac{1}{2} \int_{-20}^{20} (20^2 - y^2) dy \\ &= \int_0^{20} (20^2 - y^2) dy \\ &= \left[20^2 y - \frac{y^3}{3} \right]_0^{20} \\ &= \frac{2}{3} 20^3 = \frac{16,000}{3} \end{aligned}$$


Optional — Cylindrical shells

Let us return to Example 1.6.5 in which we rotate a region around the y -axis. Here we show another solution to this problem which is obtained by slicing the region into vertical strips. When rotated about the y -axis, each such strip sweeps out a thin cylindrical shell. Hence the name of this approach (and this subsection).

✓ Example 1.6.9 Revolving yet again

The region between the curve $y = \sqrt{x}$, and the lines $y = 0$, $x = 0$ and $x = 4$ is rotated around the line $x = 0$. Find the volume of the region swept out.

Solution

- Consider the region and cut it into thin vertical strips of width dx . 
- When we rotate the region about the line $y = 0$, each strip sweeps out a thin cylindrical shell
 - whose radius is x ,
 - height is \sqrt{x} , and
 - thickness is dx and hence
 - has volume $2\pi \times \text{radius} \times \text{height} \times \text{thickness} = 2\pi x^{3/2} dx$.
- As our leftmost strip is at $x = 0$ and our rightmost strip is at $x = 4$, the total

$$\text{Volume} = \int_0^4 2\pi x^{3/2} dx = \left[\frac{4\pi}{5} x^{5/2} \right]_0^4 = \frac{4\pi}{5} \cdot 32 = \frac{128\pi}{5}$$

which (thankfully) agrees with our previous computation.

Exercises

Stage 1

? 1

Consider a right circular cone.



What shape are horizontal cross-sections? Are the vertical cross-sections the same?

? 2

Two potters start with a block of clay h units tall, and identical square cookie cutters. They form columns by pushing the square cookie cutter straight down over the clay, so that its cross-section is the same square as the cookie cutter. Potter A pushes their cookie cutter down while their clay block is sitting motionless on a table; Potter B pushes their cookie cutter down while their clay block is rotating on a potter's wheel, so their column looks twisted. Which column has greater volume?



? 3

Let R be the region bounded above by the graph of $y = f(x)$ shown below and bounded below by the x -axis, from $x = 0$ to $x = 6$. Sketch the washers that are formed by rotating R about the y -axis. In your sketch, label all the radii in terms of y , and label the thickness.



? 4 (*)

Write down definite integrals that represent the following quantities. *Do not evaluate the integrals explicitly.*

1. The volume of the solid obtained by rotating around the x -axis the region between the x -axis and $y = \sqrt{x} e^{x^2}$ for $0 \leq x \leq 3$.
2. The volume of the solid obtained by revolving the region bounded by the curves $y = x^2$ and $y = x + 2$ about the line $x = 3$.

? 5 (*)

Write down definite integrals that represent the following quantities. *Do not evaluate the integrals explicitly.*

1. The volume of the solid obtained by rotating the finite plane region bounded by the curves $y = 1 - x^2$ and $y = 4 - 4x^2$ about the line $y = -1$.
2. The volume of the solid obtained by rotating the finite plane region bounded by the curve $y = x^2 - 1$ and the line $y = 0$ about the line $x = 5$.

? 6 (*)

Write down a definite integral that represents the volume of the solid obtained by rotating around the line $y = -1$ the region between the curves $y = x^2$ and $y = 8 - x^2$. *Do not evaluate the integrals explicitly.*

? 7

A tetrahedron is a three-dimensional shape with four faces, each of which is an equilateral triangle. (You might have seen this shape as a 4-sided die; think of a pyramid with a triangular base.) Using the methods from this section, calculate the volume of a tetrahedron with side-length ℓ . You may assume without proof that the height of a tetrahedron with side-length ℓ is $\sqrt{\frac{2}{3}}\ell$.



Stage 2

? 8 (*)

Let $a > 0$ be a constant. Let R be the finite region bounded by the graph of $y = 1 + \sqrt{x}e^{x^2}$, the line $y = 1$, and the line $x = a$. Using vertical slices, find the volume generated when R is rotated about the line $y = 1$.

? 9 (*)

Find the volume of the solid generated by rotating the finite region bounded by $y = 1/x$ and $3x + 3y = 10$ about the x -axis.

? 10 (*)

Let R be the region inside the circle $x^2 + (y - 2)^2 = 1$. Let S be the solid obtained by rotating R about the x -axis.

1. Write down an integral representing the volume of S .
2. Evaluate the integral you wrote down in part (a).

? 11 (*)

The region R is the portion of the first quadrant which is below the parabola $y^2 = 8x$ and above the hyperbola $y^2 - x^2 = 15$.

1. Sketch the region R .
2. Find the volume of the solid obtained by revolving R about the x axis.

? 12 (*)

The region R is bounded by $y = \log x$, $y = 0$, $x = 1$ and $x = 2$. (Recall that we are using $\log x$ to denote the logarithm of x with base e . In other courses it is often denoted $\ln x$.)

1. Sketch the region R .
2. Find the volume of the solid obtained by revolving this region about the y axis.

? 13 (*)

The finite region between the curves $y = \cos(\frac{x}{2})$ and $y = x^2 - \pi^2$ is rotated about the line $y = -\pi^2$. Using vertical slices (disks and/or washers), find the volume of the resulting solid.

? 14 (*)

The solid V is 2 meters high and has square horizontal cross sections. The length of the side of the square cross section at height x meters above the base is $\frac{2}{1+x}$ m. Find the volume of this solid.

? 15 (*)

Consider a solid whose base is the finite portion of the xy -plane bounded by the curves $y = x^2$ and $y = 8 - x^2$. The cross-sections perpendicular to the x -axis are squares with one side in the xy -plane. Compute the volume of this solid.

? 16 (*)

A frustrum of a right circular cone (as shown below) has height h . Its base is a circular disc with radius 4 and its top is a circular disc with radius 2. Calculate the volume of the frustrum.



Stage 3

? 17

The shape of the earth is often approximated by an oblate spheroid, rather than a sphere. An *oblate spheroid* is formed by rotating an ellipse about its *minor axis* (its shortest diameter).

1. Find the volume of the oblate spheroid obtained by rotating the upper (positive) half of the ellipse $(ax)^2 + (by)^2 = 1$ about the x -axis, where a and b are positive constants with $a \geq b$.
2. Suppose ¹² the earth has radius at the equator of 6378.137 km, and radius at the poles of 6356.752 km. If we model the earth as an oblate spheroid formed by rotating the upper half of the ellipse $(ax)^2 + (by)^2 = 1$ about the x -axis, what are a and b ?
3. What is the volume of this model of the earth? (Use a calculator.)
4. Suppose we had calculated the volume of the earth by modelling it as a sphere with radius 6378.137km. What would our absolute and relative errors be, compared to our oblate spheroid calculation?

? 18 (*)

Let R be the bounded region that lies between the curve $y = 4 - (x - 1)^2$ and the line $y = x + 1$.

1. Sketch R and find its area.
2. Write down a definite integral giving the volume of the region obtained by rotating R about the line $y = 5$. Do not evaluate this integral.

? 19 (*)

Let $\mathcal{R} = \{(x, y) : (x - 1)^2 + y^2 \leq 1 \text{ and } x^2 + (y - 1)^2 \leq 1\}$.

1. Sketch \mathcal{R} and find its area.
2. If \mathcal{R} rotates around the y -axis, what volume is generated?

? 20 (*)

Let \mathcal{R} be the plane region bounded by $x = 0$, $x = 1$, $y = 0$ and $y = c\sqrt{1 + x^2}$, where c is a positive constant.

1. Find the volume V_1 of the solid obtained by revolving \mathcal{R} about the x -axis.
2. Find the volume V_2 of the solid obtained by revolving \mathcal{R} about the y -axis.
3. If $V_1 = V_2$, what is the value of c ?

? 21 (*)

The graph below shows the region between $y = 4 + \pi \sin x$ and $y = 4 + 2\pi - 2x$.



The region is rotated about the line $y = -1$. Express in terms of definite integrals the volume of the resulting solid. Do not evaluate the integrals.

? 22

On a particular, highly homogeneous ¹³ planet, we observe that the density of the atmosphere h kilometres above the surface is given by the equation $\rho(h) = c2^{-h/6} \frac{\text{kg}}{\text{m}^3}$, where c is the density on the planet's surface.

1. What is the mass of the atmosphere contained in a vertical column with radius one metre, sixty kilometres high?
2. What height should a column be to contain $\frac{3000c\pi}{\log 2}$ kilograms of air?

1. Well — arguably the idea isn't too complicated and is a continuation of the idea used to compute areas in the previous section. In practice this can be quite tricky as we shall see.
2. Typically such integrals (and more) are covered in a third calculus course.
3. At least the bits of it involving integrals.
4. The first octant is the set of all points (x, y, z) with $x \geq 0$, $y \geq 0$ and $z \geq 0$.
5. Yet again what we really do is pick a natural number n , slice the octant of the sphere into n pancakes each of thickness $\Delta x = \frac{r}{n}$ and then take the limit $n \rightarrow \infty$. In the integral Δx is replaced by dx . Knowing that this is what is going to happen, we again just skip a few steps.
6. Okay okay... We missed the pupil. I'm sure there is a pun in there somewhere.
7. The method is not a core part of the course and should be considered optional.
8. Note that this is the full pancake, not just the part in the first octant.
9. We are again using our Riemann sum avoiding trick.
10. Handy things to have (when combined with cloth napkins) if your parents are coming to dinner and you want to convince them that you are "taking care of yourself".
11. A good question to ask to distract your parents from the fact you are serving frozen burritos.
12. [Earth Fact Sheet](#), NASA, accessed 2 July 2017

13. This is clearly a simplified model: air density changes all the time, and depends on lots of complicated factors aside from altitude. However, the equation we're using is not so far off from an idealized model of the earth's atmosphere, taken from [Pressure and the Gas Laws](#) by H.P. Schmid, accessed 3 July 2017.

This page titled [1.6: Volumes](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.7: Integration by parts

The fundamental theorem of calculus tells us that it is very easy to integrate a derivative. In particular, we know that

$$\int \frac{d}{dx}(F(x)) dx = F(x) + C$$

We can exploit this in order to develop another rule for integration — in particular a rule to help us integrate products of simpler function such as

$$\int x e^x dx$$

In so doing we will arrive at a method called “integration by parts”.

To do this we start with the product rule and integrate. Recall that the product rule says

$$\frac{d}{dx}u(x)v(x) = u'(x)v(x) + u(x)v'(x)$$

Integrating this gives

$$\begin{aligned} \int [u'(x)v(x) + u(x)v'(x)] dx &= [\text{a function whose derivative is } u'v + uv'] + C \\ &= u(x)v(x) + C \end{aligned}$$

Now this, by itself, is not terribly useful. In order to apply it we need to have a function whose integrand is a sum of products that is in exactly this form $u'(x)v(x) + u(x)v'(x)$. This is far too specialised.

However if we tease this apart a little:

$$\int [u'(x)v(x) + u(x)v'(x)] dx = \int u'(x)v(x) dx + \int u(x)v'(x) dx$$

Bring one of the integrals to the left-hand side

$$u(x)v(x) - \int u'(x)v(x) dx = \int u(x)v'(x) dx$$

Swap left and right sides

$$\int u(x)v'(x) dx = u(x)v(x) - \int u'(x)v(x) dx$$

In this form we take the integral of one product and express it in terms of the integral of a different product. If we express it like that, it doesn't seem too useful. However, if the second integral is easier, then this process helps us.

Let us do a simple example before explaining this more generally.

✓ Example 1.7.1 $\int x e^x dx$

Compute the integral $\int x e^x dx$.

Solution

- We start by taking the equation above

$$\int u(x)v'(x) dx = u(x)v(x) - \int u'(x)v(x) dx$$

- Now set $u(x) = x$ and $v'(x) = e^x$. How did we know how to make this choice? We will explain some strategies later. For now, let us just accept this choice and keep going.
- In order to use the formula we need to know $u'(x)$ and $v(x)$. In this case it is quite straightforward: $u'(x) = 1$ and $v(x) = e^x$.

- Plug everything into the formula:

$$\int x e^x dx = x e^x - \int e^x dx$$

So our original more difficult integral has been turned into a question of computing an easy one.

$$= x e^x - e^x + C$$

- We can check our answer by differentiating:

$$\begin{aligned} \frac{d}{dx}(x e^x - e^x + C) &= \underbrace{x e^x + 1 \cdot e^x}_{\text{by product rule}} - e^x + 0 \\ &= x e^x \end{aligned} \quad \text{as required.}$$

The process we have used in the above example is called “integration by parts”. When our integrand is a product we try to write it as $u(x)v'(x)$ — we need to choose one factor to be $u(x)$ and the other to be $v'(x)$. We then compute $u'(x)$ and $v(x)$ and then apply the following theorem:

Theorem 1.7.2 Integration by parts

Let $u(x)$ and $v(x)$ be continuously differentiable. Then

$$\int u(x) v'(x) dx = u(x) v(x) - \int v(x) u'(x) dx$$

If we write dv for $v'(x) dx$ and du for $u'(x) dx$ (as the substitution rule suggests), then the formula becomes

$$\int u dv = u v - \int v du$$

The application of this formula is known as integration by parts.

The corresponding statement for definite integrals is

$$\int_a^b u(x) v'(x) dx = u(b) v(b) - u(a) v(a) - \int_a^b v(x) u'(x) dx$$

Integration by parts is not as easy to apply as the product rule for derivatives. This is because it relies on us

1. judiciously choosing $u(x)$ and $v'(x)$, then
2. computing $u'(x)$ and $v(x)$ — which requires us to antidifferentiate $v'(x)$, and finally
3. that the integral $\int u'(x)v(x) dx$ is easier than the integral we started with.

Notice that any antiderivative of $v'(x)$ will do. All antiderivatives of $v'(x)$ are of the form $v(x) + A$ with A a constant. Putting this into the integration by parts formula gives

$$\begin{aligned} \int u(x) v'(x) dx &= u(x) (v(x) + A) - \int u'(x) (v(x) + A) dx \\ &= u(x)v(x) + Au(x) - \int u'(x)v(x) dx - A \underbrace{\int u'(x) dx}_{=Au(x)+C} \\ &= u(x)v(x) - \int u'(x)v(x) dx + C \end{aligned}$$

So that constant A will always cancel out.

In most applications (but not all) our integrand will be a product of two factors so we have two choices for $u(x)$ and $v'(x)$. Typically one of these choices will be “good” (in that it results in a simpler integral) while the other will be “bad” (we cannot antidifferentiate our choice of $v'(x)$ or the resulting integral is harder). Let us illustrate what we mean by returning to our previous example.

✓ Example 1.7.3 $\int x e^x dx$ — again

Our integrand is the product of two factors

$$x \quad \text{and} \quad e^x$$

This gives us two obvious choices of u and v' :

$$u(x) = x \quad v'(x) = e^x$$

or

$$u(x) = e^x \quad v'(x) = x$$

We should explore both choices:

1. If take $u(x) = x$ and $v'(x) = e^x$. We then quickly compute

$$u'(x) = 1 \quad \text{and} \quad v(x) = e^x$$

which means we will need to integrate (in the right-hand side of the integration by parts formula)

$$\int u'(x)v(x) dx = \int 1 \cdot e^x dx$$

which looks straightforward. This is a good indication that this is the right choice of $u(x)$ and $v'(x)$.

2. But before we do that, we should also explore the other choice, namely $u(x) = e^x$ and $v'(x) = x$. This implies that

$$u'(x) = e^x \quad \text{and} \quad v(x) = \frac{1}{2}x^2$$

which means we need to integrate

$$\int u'(x)v(x) dx = \int \frac{1}{2}x^2 \cdot e^x dx.$$

This is at least as hard as the integral we started with. Hence we should try the first choice.

With our choice made, we integrate by parts to get

$$\begin{aligned} \int x e^x dx &= x e^x - \int e^x dx \\ &= x e^x - e^x + C. \end{aligned}$$

The above reasoning is a very typical workflow when using integration by parts.

Integration by parts is often used

- to eliminate factors of x from an integrand like $x e^x$ by using that $\frac{d}{dx} x = 1$ and
- to eliminate a $\log x$ from an integrand by using that $\frac{d}{dx} \log x = \frac{1}{x}$ and
- to eliminate inverse trig functions, like $\arctan x$, from an integrand by using that, for example, $\frac{d}{dx} \arctan x = \frac{1}{1+x^2}$.

✓ Example 1.7.4 $\int x \sin x dx$

Solution

- Again we have a product of two factors giving us two possible choices.

1. If we choose $u(x) = x$ and $v'(x) = \sin x$, then we get

$$u'(x) = 1 \quad \text{and} \quad v(x) = -\cos x$$

which is looking promising.

2. On the other hand if we choose $u(x) = \sin x$ and $v'(x) = x$, then we have

$$u'(x) = \cos x \quad \text{and} \quad v(x) = \frac{1}{2}x^2$$

which is looking worse — we'd need to integrate $\int \frac{1}{2}x^2 \cos x \, dx$.

- So we stick with the first choice. Plugging $u(x) = x, v(x) = -\cos x$ into integration by parts gives us

$$\begin{aligned}\int x \sin x \, dx &= -x \cos x - \int 1 \cdot (-\cos x) \, dx \\ &= -x \cos x + \sin x + C\end{aligned}$$

- Again we can check our answer by differentiating:

$$\begin{aligned}\frac{d}{dx}(-x \cos x + \sin x + C) &= -\cos x + x \sin x + \cos x + 0 \\ &= x \sin x \checkmark\end{aligned}$$

Once we have practised this a bit we do not really need to write as much. Let us solve it again, but showing only what we need to.

Solution

- We use integration by parts to solve the integral.
- Set $u(x) = x$ and $v'(x) = \sin x$. Then $u'(x) = 1$ and $v(x) = -\cos x$, and

$$\begin{aligned}\int x \sin x \, dx &= -x \cos x + \int \cos x \, dx \\ &= -x \cos x + \sin x + C.\end{aligned}$$

It is pretty standard practice to reduce the notation even further in these problems. As noted above, many people write the integration by parts formula as

$$\int u \, dv = uv - \int v \, du$$

where du, dv are shorthand for $u'(x) \, dx, v'(x) \, dx$. Let us write up the previous example using this notation.

✓ Example 1.7.5 $\int x \sin x \, dx$ yet again

Solution

Using integration by parts, we set $u = x$ and $dv = \sin x \, dx$. This makes $du = 1 \, dx$ and $v = -\cos x$. Consequently

$$\begin{aligned}\int x \sin x \, dx &= \int u \, dv \\ &= uv - \int v \, du \\ &= -x \cos x + \int \cos x \, dx \\ &= -x \cos x + \sin x + C\end{aligned}$$

You can see that this is a very neat way to write up these problems and we will continue using this shorthand in the examples that follow below.

We can also use integration by parts to eliminate higher powers of x . We just need to apply the method more than once.

✓ Example 1.7.6 $\int x^2 e^x \, dx$

Solution

- Let $u = x^2$ and $dv = e^x \, dx$. This then gives $du = 2x \, dx$ and $v = e^x$, and

$$\int x^2 e^x \, dx = x^2 e^x - \int 2x e^x \, dx$$

- So we have reduced the problem of computing the original integral to one of integrating $2xe^x$. We know how to do this — just integrate by parts again:

$$\begin{aligned} \int x^2 e^x dx &= x^2 e^x - \int 2xe^x dx && \text{set } u = 2x, dv = e^x dx \\ &= x^2 e^x - \left(2xe^x - \int 2e^x dx \right) && \text{since } du = 2 dx, v = e^x \\ &= x^2 e^x - 2xe^x + 2e^x + C \end{aligned}$$

- We can, if needed, check our answer by differentiating:

$$\begin{aligned} \frac{d}{dx} (x^2 e^x - 2xe^x + 2e^x + C) \\ &= (x^2 e^x + 2xe^x) - (2xe^x + 2e^x) + 2e^x + 0 \\ &= x^2 e^x \checkmark \end{aligned}$$

A similar iterated application of integration by parts will work for integrals

$$\int P(x) (Ae^{ax} + B \sin(bx) + C \cos(cx)) dx$$

where $P(x)$ is a polynomial and A, B, C, a, b, c are constants.

Now let us look at integrands containing logarithms. We don't know the antiderivative of $\log x$, but we can eliminate $\log x$ from an integrand by using integration by parts with $u = \log x$. Remember $\log x = \log_e x = \ln x$.

✓ Example 1.7.7 $\int x \log x dx$

Solution

- We have two choices for u and dv .
 1. Set $u = x$ and $dv = \log x dx$. This gives $du = dx$ but v is hard to compute — we haven't done it yet¹. Before we go further along this path, we should look to see what happens with the other choice.
 2. Set $u = \log x$ and $dv = x dx$. This gives $du = \frac{1}{x} dx$ and $v = \frac{1}{2}x^2$, and we have to integrate

$$\int v du = \int \frac{1}{x} \cdot \frac{1}{2}x^2 dx$$

which is easy.

- So we proceed with the second choice.

$$\begin{aligned} \int x \log x dx &= \frac{1}{2}x^2 \log x - \int \frac{1}{2}x dx \\ &= \frac{1}{2}x^2 \log x - \frac{1}{4}x^2 + C \end{aligned}$$

- We can check our answer quickly:

$$\frac{d}{dx} \left(\frac{x^2}{2} \log x - \frac{x^2}{4} + C \right) = x \log x + \frac{x^2}{2} \frac{1}{x} - \frac{x}{2} + 0 = x \log x$$

✓ Example 1.7.8 $\int \log x dx$

It is not immediately obvious that one should use integration by parts to compute the integral

$$\int \log x dx$$

since the integrand is not a product. But we should persevere — indeed this is a situation where our shorter notation helps to clarify how to proceed.

Solution

- In the previous example we saw that we could remove the factor $\log x$ by setting $u = \log x$ and using integration by parts. Let us try repeating this. When we make this choice, we are then forced to take $dv = dx$ — that is we choose $v'(x) = 1$. Once we have made this sneaky move everything follows quite directly.
- We then have $du = \frac{1}{x} dx$ and $v = x$, and the integration by parts formula gives us

$$\begin{aligned} \int \log x \, dx &= x \log x - \int \frac{1}{x} \cdot x \, dx \\ &= x \log x - \int 1 \, dx \\ &= x \log x - x + C \end{aligned}$$

- As always, it is a good idea to check our result by verifying that the derivative of the answer really is the integrand.

$$\frac{d}{dx}(x \ln x - x + C) = \ln x + x \cdot \frac{1}{x} - 1 + 0 = \ln x$$

The same method works almost exactly to compute the antiderivatives of $\arcsin(x)$ and $\arctan(x)$:

✓ Example 1.7.9 $\int \arctan(x) \, dx$ and $\int \arcsin(x) \, dx$

Compute the antiderivatives of the inverse sine and inverse tangent functions.

Solution

- Again neither of these integrands are products, but that is no impediment. In both cases we set $dv = dx$ (ie $v'(x) = 1$) and choose $v(x) = x$.
- For inverse tan we choose $u = \arctan(x)$, so $du = \frac{1}{1+x^2} dx$:

$$\int \arctan(x) \, dx = x \arctan(x) - \int x \cdot \frac{1}{1+x^2} \, dx$$

now use substitution rule with $w(x) = 1 + x^2$, $w'(x) = 2x$

$$\begin{aligned} &= x \arctan(x) - \int \frac{w'(x)}{2} \cdot \frac{1}{w} \, dx \\ &= x \arctan(x) - \frac{1}{2} \int \frac{1}{w} \, dw \\ &= x \arctan(x) - \frac{1}{2} \log |w| + C \\ &= x \arctan(x) - \frac{1}{2} \log |1 + x^2| + C \quad \text{but } 1 + x^2 > 0, \text{ so} \\ &= x \arctan(x) - \frac{1}{2} \log(1 + x^2) + C \end{aligned}$$

- Similarly for inverse sine we choose $u = \arcsin(x)$ so $du = \frac{1}{\sqrt{1-x^2}} dx$:

$$\int \arcsin(x) \, dx = x \arcsin(x) - \int \frac{x}{\sqrt{1-x^2}} \, dx$$

Now use substitution rule with $w(x) = 1 - x^2$, $w'(x) = -2x$

$$\begin{aligned} &= x \arcsin(x) - \int \frac{-w'(x)}{2} \cdot w^{-1/2} \, dx \\ &= x \arcsin(x) + \frac{1}{2} \int w^{-1/2} \, dw \\ &= x \arcsin(x) + \frac{1}{2} \cdot 2w^{1/2} + C \\ &= x \arcsin(x) + \sqrt{1-x^2} + C \end{aligned}$$

- Both can be checked quite quickly by differentiating — but we leave that as an exercise for the reader.

There are many other examples we could do, but we'll finish with a tricky one.

✓ Example 1.7.10 $\int e^x \sin x \, dx$

Solution

Let us attempt this one a little naively and then we'll come back and do it more carefully (and successfully).

- We can choose either $u = e^x$, $dv = \sin x \, dx$ or the other way around.
 1. Let $u = e^x$, $dv = \sin x \, dx$. Then $du = e^x \, dx$ and $v = -\cos x$. This gives

$$\int e^x \sin x = -e^x \cos x + \int e^x \cos x \, dx$$

So we are left with an integrand that is very similar to the one we started with. What about the other choice?

2. Let $u = \sin x$, $dv = e^x \, dx$. Then $du = \cos x \, dx$ and $v = e^x$. This gives

$$\int e^x \sin x = e^x \sin x - \int e^x \cos x \, dx$$

So we are again left with an integrand that is very similar to the one we started with.

- How do we proceed? — It turns out to be easier if you do both $\int e^x \sin x \, dx$ and $\int e^x \cos x \, dx$ simultaneously. We do so in the next example.

✓ Example 1.7.11 $\int_a^b e^x \sin x \, dx$ and $\int_a^b e^x \cos x \, dx$

This time we're going to do the two integrals

$$I_1 = \int_a^b e^x \sin x \, dx \quad I_2 = \int_a^b e^x \cos x \, dx$$

at more or less the same time.

- First

$$I_1 = \int_a^b e^x \sin x \, dx = \int_a^b u \, dv$$

Choose $u = e^x$, $dv = \sin x \, dx$, so $v = -\cos x$, $du = e^x \, dx$

$$= \left[-e^x \cos x \right]_a^b + \int_a^b e^x \cos x \, dx$$

We have not found I_1 but we have related it to I_2 .

$$I_1 = \left[-e^x \cos x \right]_a^b + I_2$$

- Now start over with I_2 .

$$I_2 = \int_a^b e^x \cos x \, dx = \int_a^b u \, dv$$

Choose $u = e^x$, $dv = \cos x \, dx$, so $v = \sin x$, $du = e^x \, dx$

$$= \left[e^x \sin x \right]_a^b - \int_a^b e^x \sin x \, dx$$

Once again, we have not found I_2 but we have related it back to I_1 .

$$I_2 = \left[e^x \sin x \right]_a^b - I_1$$

- So summarising, we have

$$I_1 = \left[-e^x \cos x \right]_a^b + I_2 \quad I_2 = \left[e^x \sin x \right]_a^b - I_1$$

- So now, substitute the expression for I_2 from the second equation into the first equation to get

$$I_1 = \left[-e^x \cos x + e^x \sin x \right]_a^b - I_1$$

which implies $I_1 = \frac{1}{2} \left[e^x (\sin x - \cos x) \right]_a^b$

If we substitute the other way around we get

$$I_2 = \left[e^x \sin x + e^x \cos x \right]_a^b - I_2$$

which implies $I_2 = \frac{1}{2} \left[e^x (\sin x + \cos x) \right]_a^b$

That is,

$$\int_a^b e^x \sin x \, dx = \frac{1}{2} \left[e^x (\sin x - \cos x) \right]_a^b$$

$$\int_a^b e^x \cos x \, dx = \frac{1}{2} \left[e^x (\sin x + \cos x) \right]_a^b$$

- This also says, for example, that $\frac{1}{2}e^x(\sin x - \cos x)$ is an antiderivative of $e^x \sin x$ so that

$$\int e^x \sin x \, dx = \frac{1}{2} e^x (\sin x - \cos x) + C$$

- Note that we can always check whether or not this is correct. It is correct if and only if the derivative of the right hand side is $e^x \sin x$. Here goes. By the product rule

$$= \frac{d}{dx} \left[\frac{1}{2} e^x (\sin x - \cos x) + C \right]$$

$$= \frac{1}{2} \left[e^x (\sin x - \cos x) + e^x (\cos x + \sin x) \right] = e^x \sin x$$

which is the desired derivative.

- There is another way to find $\int e^x \sin x \, dx$ and $\int e^x \cos x \, dx$ that, in contrast to the above computations, doesn't involve any trickery. But it does require the use of complex numbers and so is beyond the scope of this course. The secret is to use that $\sin x = \frac{e^{ix} - e^{-ix}}{2i}$ and $\cos x = \frac{e^{ix} + e^{-ix}}{2}$, where i is the square root of -1 of the complex number system. See Example B.2.6.

Exercises

Stage 1

? 1

The method of integration by substitution comes from the _____ rule for differentiation.

The method of integration by parts comes from the _____ rule for differentiation.

? 2

Suppose you want to evaluate an integral using integration by parts. You choose part of your integrand to be u , and part to be dv . The part chosen as u will be: (differentiated, antiderivated). The part chosen as dv will be: (differentiated, antiderivated).

? 3

Let $f(x)$ and $g(x)$ be differentiable functions. Using the quotient rule for differentiation, give an equivalent expression to $\int \frac{f'(x)}{g(x)} dx$.

? 4

Suppose we want to use integration by parts to evaluate $\int u(x) \cdot v'(x) dx$ for some differentiable functions u and v . We need to find an antiderivative of $v'(x)$, but there are infinitely many choices. Show that every antiderivative of $v'(x)$ gives an equivalent final answer.

? 5

Suppose you want to evaluate $\int f(x) dx$ using integration by parts. Explain why $dv = f(x) dx$, $u = 1$ is generally a bad choice.

Note: compare this to Example 1.7.8, where we chose $u = f(x)$, $dv = 1 dx$.

Stage 2

? 6 (*)

Evaluate $\int x \log x dx$.

? 7 (*)

Evaluate $\int \frac{\log x}{x^7} dx$.

? 8 (*)

Evaluate $\int_0^\pi x \sin x dx$.

? 9 (*)

Evaluate $\int_0^{\frac{\pi}{2}} x \cos x dx$.

? 10

Evaluate $\int x^3 e^x dx$.

? 11

Evaluate $\int x \log^3 x \, dx$.

? 12

Evaluate $\int x^2 \sin x \, dx$.

? 13

Evaluate $\int (3t^2 - 5t + 6) \log t \, dt$.

? 14

Evaluate $\int \sqrt{s} e^{\sqrt{s}} \, ds$.

? 15

Evaluate $\int \log^2 x \, dx$.

? 16

Evaluate $\int 2x e^{x^2+1} \, dx$.

? 17 (*)

Evaluate $\int \arccos y \, dy$.

Stage 3

? 18 (*)

Evaluate $\int 4y \arctan(2y) \, dy$.

? 19

Evaluate $\int x^2 \arctan x \, dx$.

? 20

Evaluate $\int e^{x/2} \cos(2x) \, dx$.

? 21

Evaluate $\int \sin(\log x) dx$.

? 22

Evaluate $\int 2^{x+\log_2 x} dx$.

? 23

Evaluate $\int e^{\cos x} \sin(2x) dx$.

? 24

Evaluate $\int \frac{xe^{-x}}{(1-x)^2} dx$.

? 25 (*)

A reduction formula.

1. Derive the reduction formula

$$\int \sin^n(x) dx = -\frac{\sin^{n-1}(x) \cos(x)}{n} + \frac{n-1}{n} \int \sin^{n-2}(x) dx.$$

2. Calculate $\int_0^{\pi/2} \sin^8(x) dx$.

? 26 (*)

Let R be the part of the first quadrant that lies below the curve $y = \arctan x$ and between the lines $x = 0$ and $x = 1$.

1. Sketch the region R and determine its area.
2. Find the volume of the solid obtained by rotating R about the y -axis.

? 27 (*)

Let R be the region between the curves $T(x) = \sqrt{x}e^{3x}$ and $B(x) = \sqrt{x}(1+2x)$ on the interval $0 \leq x \leq 3$. (It is true that $T(x) \geq B(x)$ for all $0 \leq x \leq 3$.) Compute the volume of the solid formed by rotating R about the x -axis.

? 28 (*)

Let $f(0) = 1$, $f(2) = 3$ and $f'(2) = 4$. Calculate $\int_0^4 f''(\sqrt{x}) dx$.

? 29

Evaluate $\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{2}{n} \left(\frac{2}{n}i - 1 \right) e^{\frac{2}{n}i-1}$.

1. We will soon.

This page titled [1.7: Integration by parts](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.8: Trigonometric Integrals

Integrals of polynomials of the trigonometric functions $\sin x$, $\cos x$, $\tan x$ and so on, are generally evaluated by using a combination of simple substitutions and trigonometric identities. There are of course a very large number ¹ of trigonometric identities, but usually we use only a handful of them. The most important three are:

 Equation 1.8.1

$$\sin^2 x + \cos^2 x = 1$$

 Equation 1.8.2

$$\sin(2x) = 2 \sin x \cos x$$

 Equation 1.8.3

$$\begin{aligned} \cos(2x) &= \cos^2 x - \sin^2 x \\ &= 2 \cos^2 x - 1 \\ &= 1 - 2 \sin^2 x \end{aligned}$$

Notice that the last two lines of Equation 1.8.3 follow from the first line by replacing either $\sin^2 x$ or $\cos^2 x$ using Equation 1.8.1. It is also useful to rewrite these last two lines:

 Equation 1.8.4

$$\sin^2 x = \frac{1 - \cos(2x)}{2}$$

 Equation 1.8.5

$$\cos^2 x = \frac{1 + \cos(2x)}{2}$$

These last two are particularly useful since they allow us to rewrite higher powers of sine and cosine in terms of lower powers. For example:

$$\begin{aligned} \sin^4(x) &= \left[\frac{1 - \cos(2x)}{2} \right]^2 && \text{by Equation 1.8.4} \\ &= \frac{1}{4} - \frac{1}{2} \cos(2x) + \frac{1}{4} \underbrace{\cos^2(2x)}_{\text{do it again}} && \text{use Equation 1.8.5} \\ &= \frac{1}{4} - \frac{1}{2} \cos(2x) + \frac{1}{8} (1 + \cos(4x)) \\ &= \frac{3}{8} - \frac{1}{2} \cos(2x) + \frac{1}{8} \cos(4x) \end{aligned}$$

So while it was hard to integrate $\sin^4(x)$ directly, the final expression is quite straightforward (with a little substitution rule).

There are many such tricks for integrating powers of trigonometric functions. Here we concentrate on two families

$$\int \sin^m x \cos^n x \, dx \quad \text{and} \quad \int \tan^m x \sec^n x \, dx$$

for integer n, m . The details of the technique depend on the parity of n and m — that is, whether n and m are even or odd numbers.

Integrating $\int \sin^m x \cos^n x \, dx$

One of n and m is odd

Consider the integral $\int \sin^2 x \cos x \, dx$. We can integrate this by substituting $u = \sin x$ and $du = \cos x \, dx$. This gives

$$\begin{aligned}\int \sin^2 x \cos x \, dx &= \int u^2 \, du \\ &= \frac{1}{3} u^3 + C = \frac{1}{3} \sin^3 x + C\end{aligned}$$

This method can be used whenever n is an odd integer.

- Substitute $u = \sin x$ and $du = \cos x \, dx$.
- This leaves an even power of cosines — convert them using $\cos^2 x = 1 - \sin^2 x = 1 - u^2$.

Here is an example.

✓ Example 1.8.6 $\int \sin^2 x \cos^3 x \, dx$

Start by factoring off one power of $\cos x$ to combine with dx to get $\cos x \, dx = du$.

$$\begin{aligned}\int \sin^2 x \cos^3 x \, dx &= \int \underbrace{\sin^2 x}_{=u^2} \underbrace{\cos^2 x}_{=1-u^2} \underbrace{\cos x \, dx}_{=du} \quad \text{set } u = \sin x \\ &= \int u^2 (1 - u^2) \, du \\ &= \frac{u^3}{3} - \frac{u^5}{5} + C \\ &= \frac{\sin^3 x}{3} - \frac{\sin^5 x}{5} + C\end{aligned}$$

Of course if m is an odd integer we can use the same strategy with the roles of $\sin x$ and $\cos x$ exchanged. That is, we substitute $u = \cos x$, $du = -\sin x \, dx$ and $\sin^2 x = 1 - \cos^2 x = 1 - u^2$.

Both n and m are even

If m and n are both even, the strategy is to use the trig identities 1.8.4 and 1.8.5 to get back to the m or n odd case. This is typically more laborious than the previous case we studied. Here are a couple of examples that arise quite commonly in applications.

✓ Example 1.8.7 $\int \cos^2 x \, dx$

By 1.8.5

$$\int \cos^2 x \, dx = \frac{1}{2} \int [1 + \cos(2x)] \, dx = \frac{1}{2} \left[x + \frac{1}{2} \sin(2x) \right] + C$$

✓ Example 1.8.8 $\int \cos^4 x \, dx$

First we'll prepare the integrand $\cos^4 x$ for easy integration by applying 1.8.5 a couple times. We have already used 1.8.5 once to get

$$\cos^2 x = \frac{1}{2} [1 + \cos(2x)]$$

Squaring it gives

$$\cos^4 x = \frac{1}{4} [1 + \cos(2x)]^2 = \frac{1}{4} + \frac{1}{2} \cos(2x) + \frac{1}{4} \cos^2(2x)$$

Now by 1.8.5 a second time

$$\begin{aligned}\cos^4 x &= \frac{1}{4} + \frac{1}{2} \cos(2x) + \frac{1}{4} \frac{1 + \cos(4x)}{2} \\ &= \frac{3}{8} + \frac{1}{2} \cos(2x) + \frac{1}{8} \cos(4x)\end{aligned}$$

Now it's easy to integrate

$$\begin{aligned}\int \cos^4 x \, dx &= \frac{3}{8} \int dx + \frac{1}{2} \int \cos(2x) \, dx + \frac{1}{8} \int \cos(4x) \, dx \\ &= \frac{3}{8} x + \frac{1}{4} \sin(2x) + \frac{1}{32} \sin(4x) + C\end{aligned}$$

✓ Example 1.8.9 $\int \cos^2 x \sin^2 x dx$

Here we apply both 1.8.4 and 1.8.5.

$$\begin{aligned} \int \cos^2 x \sin^2 x dx &= \frac{1}{4} \int [1 + \cos(2x)] [1 - \cos(2x)] dx \\ &= \frac{1}{4} \int [1 - \cos^2(2x)] dx \end{aligned}$$

We can then apply 1.8.5 again

$$\begin{aligned} &= \frac{1}{4} \int \left[1 - \frac{1}{2}(1 + \cos(4x))\right] dx \\ &= \frac{1}{8} \int [1 - \cos(4x)] dx \\ &= \frac{1}{8}x - \frac{1}{32}\sin(4x) + C \end{aligned}$$

Of! We could also have done this one using 1.8.2 to write the integrand as $\sin^2(2x)$ and then used 1.8.4 to write it in terms of $\cos(4x)$.

✓ Example 1.8.10 $\int_0^\pi \cos^2 x dx$ and $\int_0^\pi \sin^2 x dx$

Of course we can compute the definite integral $\int_0^\pi \cos^2 x dx$ by using the antiderivative for $\cos^2 x$ that we found in Example 1.8.7. But here is a trickier way to evaluate that integral, and also the integral $\int_0^\pi \sin^2 x dx$ at the same time, very quickly without needing the antiderivative of Example 1.8.7.

Solution

- Observe that $\int_0^\pi \cos^2 x dx$ and $\int_0^\pi \sin^2 x dx$ are equal because they represent the same area — look at the graphs below — the darkly shaded regions in the two graphs have the same area and the lightly shaded regions in the two graphs have the same area.



- Consequently,

$$\begin{aligned} \int_0^\pi \cos^2 x dx &= \int_0^\pi \sin^2 x dx = \frac{1}{2} \left[\int_0^\pi \sin^2 x dx + \int_0^\pi \cos^2 x dx \right] \\ &= \frac{1}{2} \int_0^\pi [\sin^2 x + \cos^2 x] dx \\ &= \frac{1}{2} \int_0^\pi dx \\ &= \frac{\pi}{2} \end{aligned}$$

Integrating $\int \tan^m x \sec^n x dx$

The strategy for dealing with these integrals is similar to the strategy that we used to evaluate integrals of the form $\int \sin^m x \cos^n x dx$ and again depends on the parity of the exponents n and m . It uses ²

$$\frac{d}{dx} \tan x = \sec^2 x \quad \frac{d}{dx} \sec x = \sec x \tan x \quad 1 + \tan^2 x = \sec^2 x$$

We split the methods for integrating $\int \tan^m x \sec^n x dx$ into 5 cases which we list below. These will become much more clear after an example (or two).

- When m is odd and any n — rewrite the integrand in terms of $\sin x$ and $\cos x$:

$$\begin{aligned} \tan^m x \sec^n x dx &= \left(\frac{\sin x}{\cos x} \right)^m \left(\frac{1}{\cos x} \right)^n dx \\ &= \frac{\sin^{m-1} x}{\cos^{n+m} x} \sin x dx \end{aligned}$$

and then substitute $u = \cos x$, $du = -\sin x dx$, $\sin^2 x = 1 - \cos^2 x = 1 - u^2$. See Examples 1.8.11 and 1.8.12.

- Alternatively, if m is odd and $n \geq 1$ move one factor of $\sec x \tan x$ to the side so that you can see $\sec x \tan x dx$ in the integral, and substitute $u = \sec x$, $du = \sec x \tan x dx$ and $\tan^2 x = \sec^2 x - 1 = u^2 - 1$. See Example 1.8.13.
- If n is even with $n \geq 2$, move one factor of $\sec^2 x$ to the side so that you can see $\sec^2 x dx$ in the integral, and substitute $u = \tan x$, $du = \sec^2 x dx$ and $\sec^2 x = 1 + \tan^2 x = 1 + u^2$. See Example 1.8.14.

- When m is even and $n = 0$ — that is the integrand is just an even power of tangent — we can still use the $u = \tan x$ substitution, after using $\tan^2 x = \sec^2 x - 1$ (possibly more than once) to create a $\sec^2 x$. See Example 1.8.16.
- This leaves the case n odd and m even. There are strategies like those above for treating this case. But they are more complicated and also involve more tricks (that basically have to be memorized). Examples using them are provided in the optional section entitled “Integrating $\sec x$, $\csc x$, $\sec^3 x$ and $\csc^3 x$ ”, below. A more straight forward strategy uses another technique called “partial fractions”. We shall return to this strategy after we have learned about partial fractions. See Example 1.10.5 and 1.10.6 in Section 1.10.

m is odd — odd power of tangent

In this case we rewrite the integrand in terms of sine and cosine and then substitute $u = \cos x$, $du = -\sin x dx$.

✓ Example 1.8.11 $\int \tan x dx$

Solution

- Write the integrand $\tan x = \frac{1}{\cos x} \sin x$.
- Now substitute $u = \cos x$, $du = -\sin x dx$ just as we did in treating integrands of the form $\sin^m x \cos^n x$ with m odd.

$$\begin{aligned} \int \tan x dx &= \int \frac{1}{\cos x} \sin x dx && \text{substitute } u = \cos x \\ &= \int \frac{1}{u} \cdot (-1) du \\ &= -\log |u| + C \\ &= -\log |\cos x| + C && \text{can also write in terms of secant} \\ &= \log |\cos x|^{-1} + C = \log |\sec x| + C \end{aligned}$$

✓ Example 1.8.12 $\int \tan^3 x dx$

Solution

- Write the integrand $\tan^3 x = \frac{\sin^2 x}{\cos^3 x} \sin x$.
- Again substitute $u = \cos x$, $du = -\sin x dx$. We rewrite the remaining even powers of $\sin x$ using $\sin^2 x = 1 - \cos^2 x = 1 - u^2$.
- Hence

$$\begin{aligned} \int \tan^3 x dx &= \int \frac{\sin^2 x}{\cos^3 x} \sin x dx && \text{substitute } u = \cos x \\ &= \int \frac{1-u^2}{u^3} (-1) du \\ &= \frac{u^{-2}}{2} + \log |u| + C \\ &= \frac{1}{2 \cos^2 x} + \log |\cos x| + C && \text{rewrite in terms of secant} \\ &= \frac{1}{2} \sec^2 x - \log |\sec x| + C \end{aligned}$$

m is odd and $n \geq 1$ — odd power of tangent and at least one secant

Here we collect a factor of $\tan x \sec x$ and then substitute $u = \sec x$ and $du = \sec x \tan x dx$. We can then rewrite any remaining even powers of $\tan x$ in terms of $\sec x$ using $\tan^2 x = \sec^2 x - 1 = u^2 - 1$.

✓ Example 1.8.13 $\int \tan^3 x \sec^4 x dx$

Solution

- Start by factoring off one copy of $\sec x \tan x$ and combine it with dx to form $\sec x \tan x dx$, which will be du .
- Now substitute $u = \sec x$, $du = \sec x \tan x dx$ and $\tan^2 x = \sec^2 x - 1 = u^2 - 1$.
- This gives

$$\begin{aligned}
 \int \tan^3 x \sec^4 x \, dx &= \int \underbrace{\tan^2 x}_{u^2-1} \underbrace{\sec^3 x}_{u^3} \underbrace{\sec x \tan x \, dx}_{du} \\
 &= \int [u^2 - 1] u^3 \, du \\
 &= \frac{u^6}{6} - \frac{u^4}{4} + C \\
 &= \frac{1}{6} \sec^6 x - \frac{1}{4} \sec^4 x + C
 \end{aligned}$$

$n \geq 2$ is even — a positive even power of secant

In the previous case we substituted $u = \sec x$, while in this case we substitute $u = \tan x$. When we do this we write $du = \sec^2 x \, dx$ and then rewrite any remaining even powers of $\sec x$ as powers of $\tan x$ using $\sec^2 x = 1 + \tan^2 x = 1 + u^2$.

✓ Example 1.8.14 $\int \sec^4 x \, dx$

Solution

- Factor off one copy of $\sec^2 x$ and combine it with dx to form $\sec^2 x \, dx$, which will be du .
- Then substitute $u = \tan x$, $du = \sec^2 x \, dx$ and rewrite any remaining even powers of $\sec x$ as powers of $\tan x = u$ using $\sec^2 x = 1 + \tan^2 x = 1 + u^2$.
- This gives

$$\begin{aligned}
 \int \sec^4 x \, dx &= \int \underbrace{\sec^2 x}_{1+u^2} \underbrace{\sec^2 x \, dx}_{du} \\
 &= \int [1 + u^2] \, du \\
 &= u + \frac{u^3}{3} + C \\
 &= \tan x + \frac{1}{3} \tan^3 x + C
 \end{aligned}$$

✓ Example 1.8.15 $\int \tan^3 x \sec^4 x \, dx$ — redux

Solution

Let us revisit this example using this slightly different approach.

- Factor off one copy of $\sec^2 x$ and combine it with dx to form $\sec^2 x \, dx$, which will be du .
- Then substitute $u = \tan x$, $du = \sec^2 x \, dx$ and rewrite any remaining even powers of $\sec x$ as powers of $\tan x = u$ using $\sec^2 x = 1 + \tan^2 x = 1 + u^2$.
- This gives

$$\begin{aligned}
 \int \tan^3 x \sec^4 x \, dx &= \int \underbrace{\tan^3 x}_{u^3} \underbrace{\sec^2 x}_{1+u^2} \underbrace{\sec^2 x \, dx}_{du} \\
 &= \int [u^3 + u^5] \, du \\
 &= \frac{u^4}{4} + \frac{u^6}{6} + C \\
 &= \frac{1}{4} \tan^4 x + \frac{1}{6} \tan^6 x + C
 \end{aligned}$$

- This is not quite the same as the answer we got above in Example 1.8.13. However we can show they are (nearly) equivalent. To do so we substitute $v = \sec x$ and $\tan^2 x = \sec^2 x - 1 = v^2 - 1$:

$$\begin{aligned}
 \frac{1}{6}\tan^6 x + \frac{1}{4}\tan^4 x &= \frac{1}{6}(v^2 - 1)^3 + \frac{1}{4}(v^2 - 1)^2 \\
 &= \frac{1}{6}(v^6 - 3v^4 + 3v^2 - 1) + \frac{1}{4}(v^4 - 2v^2 + 1) \\
 &= \frac{v^6}{6} - \frac{v^4}{2} + \frac{v^2}{2} - \frac{1}{6} + \frac{v^4}{4} - \frac{v^2}{2} + \frac{1}{4} \\
 &= \frac{v^6}{6} - \frac{v^4}{4} + 0 \cdot v^2 + \left(\frac{1}{4} - \frac{1}{6}\right) \\
 &= \frac{1}{6}\sec^6 x - \frac{1}{4}\sec^4 x + \frac{1}{12}.
 \end{aligned}$$

So while $\frac{1}{6}\tan^6 x + \frac{1}{4}\tan^4 x \neq \frac{1}{6}\sec^6 x - \frac{1}{4}\sec^4 x$, they only differ by a constant. Hence both are valid antiderivatives of $\tan^3 x \sec^4 x$.

m is even and $n = 0$ — even powers of tangent

We integrate this by setting $u = \tan x$. For this to work we need to pull one factor of $\sec^2 x$ to one side to form $du = \sec^2 x dx$. To find this factor of $\sec^2 x$ we (perhaps repeatedly) apply the identity $\tan^2 x = \sec^2 x - 1$.

✓ Example 1.8.16 $\int \tan^4 x dx$

Solution

- There is no $\sec^2 x$ term present, so we try to create it from $\tan^4 x$ by using $\tan^2 x = \sec^2 x - 1$.

$$\begin{aligned}
 \tan^4 x &= \tan^2 x \cdot \tan^2 x \\
 &= \tan^2 x [\sec^2 x - 1] \\
 &= \tan^2 x \sec^2 x - \underbrace{\tan^2 x}_{\sec^2 x - 1} \\
 &= \tan^2 x \sec^2 x - \sec^2 x + 1
 \end{aligned}$$

- Now we can substitute $u = \tan x$, $du = \sec^2 x dx$.

$$\begin{aligned}
 \int \tan^4 x dx &= \int \underbrace{\tan^2 x}_{u^2} \underbrace{\sec^2 x dx}_{du} - \int \underbrace{\sec^2 x dx}_{du} + \int dx \\
 &= \int u^2 du - \int du + \int dx \\
 &= \frac{u^3}{3} - u + x + C \\
 &= \frac{\tan^3 x}{3} - \tan x + x + C
 \end{aligned}$$

✓ Example 1.8.17 $\int \tan^8 x dx$

Solution

Let us try the same approach.

- First pull out a factor of $\tan^2 x$ to create a $\sec^2 x$ factor:

$$\begin{aligned}
 \tan^8 x &= \tan^6 x \cdot \tan^2 x \\
 &= \tan^6 x \cdot [\sec^2 x - 1] \\
 &= \tan^6 x \sec^2 x - \tan^6 x
 \end{aligned}$$

The first term is now ready to be integrated, but we need to reapply the method to the second term:

$$\begin{aligned}
 &= \tan^6 x \sec^2 x - \tan^4 x \cdot [\sec^2 x - 1] \\
 &= \tan^6 x \sec^2 x - \tan^4 x \sec^2 x + \tan^4 x \quad \text{do it again} \\
 &= \tan^6 x \sec^2 x - \tan^4 x \sec^2 x + \tan^2 x \cdot [\sec^2 x - 1] \\
 &= \tan^6 x \sec^2 x - \tan^4 x \sec^2 x + \tan^2 x \sec^2 x - \tan^2 x \quad \text{and again} \\
 &= \tan^6 x \sec^2 x - \tan^4 x \sec^2 x + \tan^2 x \sec^2 x - [\sec^2 x - 1]
 \end{aligned}$$

- Hence

$$\begin{aligned}
 \int \tan^8 x \, dx &= \int [\tan^6 x \sec^2 x - \tan^4 x \sec^2 x + \tan^2 x \sec^2 x - \sec^2 x + 1] \, dx \\
 &= \int [\tan^6 x - \tan^4 x + \tan^2 x - 1] \sec^2 x \, dx + \int dx \\
 &= \int [u^6 - u^4 + u^2 - 1] \, du + x + C \\
 &= \frac{u^7}{7} - \frac{u^5}{5} + \frac{u^3}{3} - u + x + C \\
 &= \frac{1}{7} \tan^7 x - \frac{1}{5} \tan^5 x + \frac{1}{3} \tan^3 x - \tan x + x + C
 \end{aligned}$$

Indeed this example suggests that for integer $k \geq 0$:

$$\int \tan^{2k} x \, dx = \frac{1}{2k-1} \tan^{2k-1}(x) - \frac{1}{2k-3} \tan^{2k-3} x + \dots - (-1)^k \tan x + (-1)^k x + C$$

This last example also shows how we might integrate an odd power of tangent:

✓ Example 1.8.18 $\int \tan^7 x$

Solution

We follow the same steps

- Pull out a factor of $\tan^2 x$ to create a factor of $\sec^2 x$:

$$\begin{aligned}
 \tan^7 x &= \tan^5 x \cdot \tan^2 x \\
 &= \tan^5 x \cdot [\sec^2 x - 1] \\
 &= \tan^5 x \sec^2 x - \tan^5 x \quad \text{do it again} \\
 &= \tan^5 x \sec^2 x - \tan^3 x \cdot [\sec^2 x - 1] \\
 &= \tan^5 x \sec^2 x - \tan^3 x \sec^2 x + \tan^3 x \quad \text{and again} \\
 &= \tan^5 x \sec^2 x - \tan^3 x \sec^2 x + \tan x [\sec^2 x - 1] \\
 &= \tan^5 x \sec^2 x - \tan^3 x \sec^2 x + \tan x \sec^2 x - \tan x
 \end{aligned}$$

- Now we can substitute $u = \tan x$ and $du = \sec^2 x \, dx$ and also use the result from Example 1.8.11 to take care of the last term:

$$\int \tan^7 x \, dx = \int [\tan^5 x \sec^2 x - \tan^3 x \sec^2 x + \tan x \sec^2 x] \, dx - \int \tan x \, dx$$

Now factor out the common $\sec^2 x$ term and integrate $\tan x$ via Example 1.8.11

$$\begin{aligned}
 &= \int [\tan^5 x - \tan^3 x + \tan x] \sec^2 x \, dx - \log |\sec x| + C \\
 &= \int [u^5 - u^3 + u] \, du - \log |\sec x| + C \\
 &= \frac{u^6}{6} - \frac{u^4}{4} + \frac{u^2}{2} - \log |\sec x| + C \\
 &= \frac{1}{6} \tan^6 x - \frac{1}{4} \tan^4 x + \frac{1}{2} \tan^2 x - \log |\sec x| + C
 \end{aligned}$$

This example suggests that for integer $k \geq 0$:

$$\int \tan^{2k+1} x \, dx = \frac{1}{2k} \tan^{2k}(x) - \frac{1}{2k-2} \tan^{2k-2} x + \dots - (-1)^k \frac{1}{2} \tan^2 x + (-1)^k \log |\sec x| + C$$

Of course we have not considered integrals involving powers of $\cot x$ and $\csc x$. But they can be treated in much the same way as $\tan x$ and $\sec x$ were.

Optional — Integrating $\sec x$, $\csc x$, $\sec^3 x$ and $\csc^3 x$

As noted above, when n is odd and m is even, one can use similar strategies as to the previous cases. However the computations are often more involved and more tricks need to be deployed. For this reason we make this section optional — the computations are definitely non-trivial. Rather than trying to construct a coherent “method” for this case, we instead give some examples to give the idea of what to expect.

✓ Example 1.8.19 $\int \sec x \, dx$ — by trickery

Solution

There is a very sneaky trick to compute this integral.

- The standard trick for this integral is to multiply the integrand by $1 = \frac{\sec x + \tan x}{\sec x + \tan x}$

$$\sec x = \sec x \frac{\sec x + \tan x}{\sec x + \tan x} = \frac{\sec^2 x + \sec x \tan x}{\sec x + \tan x}$$

- Notice now that the numerator of this expression is exactly the derivative its denominator. Hence we can substitute $u = \sec x + \tan x$ and $du = (\sec x \tan x + \sec^2 x) \, dx$.
- Hence

$$\begin{aligned} \int \sec x \, dx &= \int \sec x \frac{\sec x + \tan x}{\sec x + \tan x} \, dx = \int \frac{\sec^2 x + \sec x \tan x}{\sec x + \tan x} \, dx \\ &= \int \frac{1}{u} \, du \\ &= \log |u| + C \\ &= \log |\sec x + \tan x| + C \end{aligned}$$

- The above trick appears both totally unguessable and very hard to remember. Fortunately, there is a simple way³ to recover the trick. Here it is.
 - The goal is to guess a function whose derivative is $\sec x$.
 - So get out a table of derivatives and look for functions whose derivatives at least contain $\sec x$. There are two:

$$\begin{aligned} \frac{d}{dx} \tan x &= \sec^2 x \\ \frac{d}{dx} \sec x &= \tan x \sec x \end{aligned}$$

- Notice that if we add these together we get

$$\begin{aligned} \frac{d}{dx} (\sec x + \tan x) &= (\sec x + \tan x) \sec x \quad \implies \\ \frac{\frac{d}{dx} (\sec x + \tan x)}{\sec x + \tan x} &= \sec x \end{aligned}$$

- We've done it! The right hand side is $\sec x$ and the left hand side is the derivative of $\log |\sec x + \tan x|$.

There is another method for integrating $\int \sec x \, dx$, that is more tedious, but more straight forward. In particular, it does not involve a memorized trick. We first use the substitution $u = \sin x$, $du = \cos x \, dx$, together with $\cos^2 x = 1 - \sin^2 x = 1 - u^2$. This converts the integral into

$$\begin{aligned} \int \sec x \, dx &= \int \frac{1}{\cos x} \, dx = \int \frac{\cos x \, dx}{\cos^2 x} \\ &= \int \frac{du}{1 - u^2} \Big|_{u=\sin x} \end{aligned}$$

The integrand $\frac{1}{1-u^2}$ is a rational function, i.e. a ratio of two polynomials. There is a procedure, called the method of partial fractions, that may be used to integrate any rational function. We shall learn about it in Section 1.10 “Partial Fractions”. The detailed evaluation of the integral $\int \sec x \, dx = \int \frac{du}{1-u^2}$ by the method of partial fractions is presented in Example 1.10.5 below.

In addition, there is a standard trick for evaluating $\int \frac{du}{1-u^2}$ that allows us to avoid going through the whole partial fractions algorithm.

✓ Example 1.8.20 $\int \sec x \, dx$ — by more trickery

Solution

We have already seen that

$$\int \sec x \, dx = \int \frac{du}{1-u^2} \Big|_{u=\sin x}$$

The trick uses the observations that

- $\frac{1}{1-u^2} = \frac{1+u-u}{1-u^2} = \frac{1}{1-u} - \frac{u}{1-u^2}$
- $\frac{1}{1-u}$ has antiderivative $-\log(1-u)$ (for $u < 1$)
- The derivative $\frac{d}{du}(1-u^2) = -2u$ of the denominator of $\frac{u}{1-u^2}$ is the same, up to a factor of -2 , as the numerator of $\frac{u}{1-u^2}$. So we can easily evaluate the integral of $\frac{u}{1-u^2}$ by substituting $v = 1-u^2$, $dv = -2u \, du$.

$$\int \frac{u \, du}{1-u^2} = \int \frac{\frac{dv}{-2}}{v} \Big|_{v=1-u^2} = -\frac{1}{2} \log(1-u^2) + C$$

Combining these observations gives

$$\begin{aligned} \int \sec x \, dx &= \left[\int \frac{du}{1-u^2} \right]_{u=\sin x} = \left[\int \frac{1}{1-u} \, du - \int \frac{u}{1-u^2} \, du \right]_{u=\sin x} \\ &= \left[-\log(1-u) + \frac{1}{2} \log(1-u^2) + C \right]_{u=\sin x} \\ &= -\log(1-\sin x) + \frac{1}{2} \log(1-\sin^2 x) + C \\ &= -\log(1-\sin x) + \frac{1}{2} \log(1-\sin x) + \frac{1}{2} \log(1+\sin x) + C \\ &= \frac{1}{2} \log \frac{1+\sin x}{1-\sin x} + C \end{aligned}$$

Example 1.8.20 has given the answer

$$\int \sec x \, dx = \frac{1}{2} \log \frac{1+\sin x}{1-\sin x} + C$$

which appears to be different than the answer in Example 1.8.19. But they are really the same since

$$\begin{aligned} \frac{1+\sin x}{1-\sin x} &= \frac{(1+\sin x)^2}{1-\sin^2 x} = \frac{(1+\sin x)^2}{\cos^2 x} \\ \implies \frac{1}{2} \log \frac{1+\sin x}{1-\sin x} &= \frac{1}{2} \log \frac{(1+\sin x)^2}{\cos^2 x} = \log \left| \frac{\sin x + 1}{\cos x} \right| = \log |\tan x + \sec x| \end{aligned}$$

Oof!

✓ Example 1.8.21 $\int \csc x \, dx$ — by the $u = \tan \frac{x}{2}$ substitution

Solution

The integral $\int \csc x \, dx$ may also be evaluated by both the methods above. That is either

- by multiplying the integrand by a cleverly chosen $1 = \frac{\cot x - \csc x}{\cot x - \csc x}$ and then substituting $u = \cot x - \csc x$, $du = (-\csc^2 x + \csc x \cot x) \, dx$, or
- by substituting $u = \cos x$, $du = -\sin x \, dx$ to give $\int \csc x \, dx = -\int \frac{du}{1-u^2}$ and then using the method of partial fractions.

These two methods give the answers

$$\int \csc x \, dx = \log |\cot x - \csc x| + C = -\frac{1}{2} \log \frac{1+\cos x}{1-\cos x} + C \quad (\star)$$

In this example, we shall evaluate $\int \csc x \, dx$ by yet a third method, which can be used to integrate rational functions ⁴A rational function of $\sin x$ and $\cos x$ is a ratio with both the numerator and denominator being finite sums of terms of the form $a \sin^m x \cos^n x$, where a is a constant and m and n are positive integers. of $\sin x$ and $\cos x$.

- This method uses the substitution

$$x = 2 \arctan u \quad \text{i.e. } u = \tan \frac{x}{2} \quad \text{and } dx = \frac{2}{1+u^2} \, du$$

— a half-angle substitution.

- To express $\sin x$ and $\cos x$ in terms of u , we first use the double angle trig identities (Equations 1.8.2 and 1.8.3 with $x \mapsto \frac{x}{2}$) to express $\sin x$ and $\cos x$ in terms of $\sin \frac{x}{2}$ and $\cos \frac{x}{2}$:

$$\begin{aligned}\sin x &= 2 \sin \frac{x}{2} \cos \frac{x}{2} \\ \cos x &= \cos^2 \frac{x}{2} - \sin^2 \frac{x}{2}\end{aligned}$$

- We then use the triangle



to express $\sin \frac{x}{2}$ and $\cos \frac{x}{2}$ in terms of u . The bottom and right hand sides of the triangle have been chosen so that $\tan \frac{x}{2} = u$. This tells us that

$$\sin \frac{x}{2} = \frac{u}{\sqrt{1+u^2}} \quad \cos \frac{x}{2} = \frac{1}{\sqrt{1+u^2}}$$

- This in turn implies that:

$$\begin{aligned}\sin x &= 2 \sin \frac{x}{2} \cos \frac{x}{2} = 2 \frac{u}{\sqrt{1+u^2}} \frac{1}{\sqrt{1+u^2}} = \frac{2u}{1+u^2} \\ \cos x &= \cos^2 \frac{x}{2} - \sin^2 \frac{x}{2} = \frac{1}{1+u^2} - \frac{u^2}{1+u^2} = \frac{1-u^2}{1+u^2}\end{aligned}$$

Oof!

- Let's use this substitution to evaluate $\int \csc x \, dx$.

$$\begin{aligned}\int \csc x \, dx &= \int \frac{1}{\sin x} \, dx = \int \frac{1+u^2}{2u} \frac{2}{1+u^2} \, du = \int \frac{1}{u} \, du \\ &= \log |u| + C = \log \left| \tan \frac{x}{2} \right| + C\end{aligned}$$

To see that this answer is really the same as that in (*), note that

$$\cot x - \csc x = \frac{\cos x - 1}{\sin x} = \frac{-2 \sin^2(x/2)}{2 \sin(x/2) \cos(x/2)} = -\tan \frac{x}{2}$$

✓ Example 1.8.22 $\int \sec^3 x \, dx$ — by trickery

Solution

The standard trick used to evaluate $\int \sec^3 x \, dx$ is integration by parts.

- Set $u = \sec x$, $dv = \sec^2 x \, dx$. Hence $du = \sec x \tan x \, dx$, $v = \tan x$ and

$$\begin{aligned}\int \sec^3 x \, dx &= \int \underbrace{\sec x}_u \underbrace{\sec^2 x \, dx}_{dv} \\ &= \underbrace{\sec x}_u \underbrace{\tan x}_v - \int \underbrace{\tan x}_v \underbrace{\sec x \tan x \, dx}_{du}\end{aligned}$$

- Since $\tan^2 x + 1 = \sec^2 x$, we have $\tan^2 x = \sec^2 x - 1$ and

$$\begin{aligned}\int \sec^3 x \, dx &= \sec x \tan x - \int [\sec^3 x - \sec x] \, dx \\ &= \sec x \tan x + \log |\sec x + \tan x| + C - \int \sec^3 x \, dx\end{aligned}$$

where we used $\int \sec x \, dx = \log |\sec x + \tan x| + C$, which we saw in Example 1.8.19.

- Now moving the $\int \sec^3 x \, dx$ from the right hand side to the left hand side

$$\begin{aligned}2 \int \sec^3 x \, dx &= \sec x \tan x + \log |\sec x + \tan x| + C \quad \text{and so} \\ \int \sec^3 x \, dx &= \frac{1}{2} \sec x \tan x + \frac{1}{2} \log |\sec x + \tan x| + C\end{aligned}$$

for a new arbitrary constant C (which is just one half the old one).

The integral $\int \sec^3 x \, dx$ can also be evaluated by two other methods.

- Substitute $u = \sin x$, $du = \cos x dx$ to convert $\int \sec^3 x dx$ into $\int \frac{du}{[1-u^2]^2}$ and evaluate the latter using the method of partial fractions. This is done in Example 1.10.6 in Section 1.10.
- Use the $u = \tan \frac{x}{2}$ substitution. We use this method to evaluate $\int \csc^3 x dx$ in Example 1.8.23, below.

✓ Example 1.8.23 $\int \csc^3 x dx$ – by the $u = \tan \frac{x}{2}$ substitution

Solution

Let us use the half-angle substitution that we introduced in Example 1.8.21.

- In this method we set

$$u = \tan \frac{x}{2} \quad dx = \frac{2}{1+u^2} du \quad \sin x = \frac{2u}{1+u^2} \quad \cos x = \frac{1-u^2}{1+u^2}$$

- The integral then becomes

$$\begin{aligned} \int \csc^3 x dx &= \int \frac{1}{\sin^3 x} dx \\ &= \int \left(\frac{1+u^2}{2u} \right)^3 \frac{2}{1+u^2} du \\ &= \frac{1}{4} \int \frac{1+2u^2+u^4}{u^3} du \\ &= \frac{1}{4} \left\{ \frac{u^{-2}}{-2} + 2 \log |u| + \frac{u^2}{2} \right\} + C \\ &= \frac{1}{8} \left\{ -\cot^2 \frac{x}{2} + 4 \log \left| \tan \frac{x}{2} \right| + \tan^2 \frac{x}{2} \right\} + C \end{aligned}$$

Oof!

- This is a perfectly acceptable answer. But if you don't like the $\frac{x}{2}$'s, they may be eliminated by using

$$\begin{aligned} \tan^2 \frac{x}{2} - \cot^2 \frac{x}{2} &= \frac{\sin^2 \frac{x}{2}}{\cos^2 \frac{x}{2}} - \frac{\cos^2 \frac{x}{2}}{\sin^2 \frac{x}{2}} \\ &= \frac{\sin^4 \frac{x}{2} - \cos^4 \frac{x}{2}}{\sin^2 \frac{x}{2} \cos^2 \frac{x}{2}} \\ &= \frac{(\sin^2 \frac{x}{2} - \cos^2 \frac{x}{2})(\sin^2 \frac{x}{2} + \cos^2 \frac{x}{2})}{\sin^2 \frac{x}{2} \cos^2 \frac{x}{2}} \\ &= \frac{\sin^2 \frac{x}{2} - \cos^2 \frac{x}{2}}{\sin^2 \frac{x}{2} \cos^2 \frac{x}{2}} \quad \text{since } \sin^2 \frac{x}{2} + \cos^2 \frac{x}{2} = 1 \\ &= \frac{-\cos x}{\frac{1}{4} \sin^2 x} \quad \text{by 1.8.2 and 1.8.3} \end{aligned}$$

and

$$\begin{aligned} \tan \frac{x}{2} &= \frac{\sin \frac{x}{2}}{\cos \frac{x}{2}} = \frac{\sin^2 \frac{x}{2}}{\sin \frac{x}{2} \cos \frac{x}{2}} \\ &= \frac{\frac{1}{2}[1 - \cos x]}{\frac{1}{2} \sin x} \end{aligned}$$

by [\knowl./knowl/eq_TRGINTtrigidentityB.html1.8.2](https://math.libretexts.org/@go/page/91712) and [\knowl./knowl/eq_TRGINTtrigidentityC.html1.8.3](https://math.libretexts.org/@go/page/91712)

So we may also write

$$\int \csc^3 x dx = -\frac{1}{2} \cot x \csc x + \frac{1}{2} \log |\csc x - \cot x| + C$$

That last optional section was a little scary — let's get back to something a little easier.

Exercises

Recall that we are using $\log x$ to denote the logarithm of x with base e . In other courses it is often denoted $\ln x$.

Stage 1**? 1**

Suppose you want to evaluate $\int_0^{\pi/4} \sin x \cos^n x \, dx$ using the substitution $u = \cos x$. Which of the following need to be true for your substitution to work?

1. n must be even
2. n must be odd
3. n must be an integer
4. n must be positive
5. n can be any real number

? 2

Evaluate $\int \sec^n x \tan x \, dx$, where n is a strictly positive integer.

? 3

Derive the identity $\tan^2 x + 1 = \sec^2 x$ from the easier-to-remember identity $\sin^2 x + \cos^2 x = 1$.

Stage 2

Questions 4 through 10 deal with powers of sines and cosines. Review Section 1.8.1 in the notes for integration strategies.

Questions 12 through 21 deal with powers of tangents and secants. Review Section 1.8.2 in the notes for strategies.

? 4 (*)

Evaluate $\int \cos^3 x \, dx$.

? 5 (*)

Evaluate $\int_0^{\pi} \cos^2 x \, dx$.

? 6 (*)

Evaluate $\int \sin^{36} t \cos^3 t \, dt$.

? 7

Evaluate $\int \frac{\sin^3 x}{\cos^4 x} \, dx$.

? 8

Evaluate $\int_0^{\pi/3} \sin^4 x \, dx$.

? 9

Evaluate $\int \sin^5 x \, dx$.

? 10

Evaluate $\int \sin^{1.2} x \cos x \, dx$.

? 11

Evaluate $\int \tan x \sec^2 x \, dx$.

? 12 (*)

Evaluate $\int \tan^3 x \sec^5 x \, dx$.

? 13 (*)

Evaluate $\int \sec^4 x \tan^{46} x \, dx$.

? 14

Evaluate $\int \tan^3 x \sec^{1.5} x \, dx$.

? 15

Evaluate $\int \tan^3 x \sec^2 x \, dx$.

? 16

Evaluate $\int \tan^4 x \sec^2 x \, dx$.

? 17

Evaluate $\int \tan^3 x \sec^{-0.7} x \, dx$.

? 18

Evaluate $\int \tan^5 x \, dx$.

? 19

Evaluate $\int_0^{\pi/6} \tan^6 x \, dx$.

? 20

Evaluate $\int_0^{\pi/4} \tan^8 x \sec^4 x \, dx$.

? 21

Evaluate $\int \tan x \sqrt{\sec x} \, dx$.

? 22

Evaluate $\int \sec^8 \theta \tan^e \theta d\theta$.

Stage 3

? 23 (*)

A reduction formula.

1. Let n be a positive integer with $n \geq 2$. Derive the reduction formula

$$\int \tan^n(x) dx = \frac{\tan^{n-1}(x)}{n-1} - \int \tan^{n-2}(x) dx.$$

2. Calculate $\int_0^{\pi/4} \tan^6(x) dx$.

? 24

Evaluate $\int \tan^5 x \cos^2 x dx$.

? 25

Evaluate $\int \frac{1}{\cos^2 \theta} d\theta$.

? 26

Evaluate $\int \cot x dx$.

? 27

Evaluate $\int e^x \sin(e^x) \cos(e^x) dx$.

? 28

Evaluate $\int \sin(\cos x) \sin^3 x dx$.

? 29

Evaluate $\int x \sin x \cos x dx$.

1. The more pedantic reader could construct an infinite list of them.
2. You will need to memorise the derivatives of tangent and secant. However there is no need to memorise $1 + \tan^2 x = \sec^2 x$. To derive it very quickly just divide $\sin^2 x + \cos^2 x = 1$ by $\cos^2 x$.
3. We thank Serban Raianu for bringing this to our attention.

This page titled [1.8: Trigonometric Integrals](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.9: Trigonometric Substitution

In this section we discuss substitutions that simplify integrals containing square roots of the form

$$\sqrt{a^2 - x^2} \quad \sqrt{a^2 + x^2} \quad \sqrt{x^2 - a^2}.$$

When the integrand contains one of these square roots, then we can use trigonometric substitutions to eliminate them. That is, we substitute

$$x = a \sin u \quad \text{or} \quad x = a \tan u \quad \text{or} \quad x = a \sec u$$

and then use trigonometric identities

$$\sin^2 \theta + \cos^2 \theta = 1 \quad \text{and} \quad 1 + \tan^2 \theta = \sec^2 \theta$$

to simplify the result. To be more precise, we can

- eliminate $\sqrt{a^2 - x^2}$ from an integrand by substituting $x = a \sin u$ to give

$$\sqrt{a^2 - x^2} = \sqrt{a^2 - a^2 \sin^2 u} = \sqrt{a^2 \cos^2 u} = |a \cos u|$$

- eliminate $\sqrt{a^2 + x^2}$ from an integrand by substituting $x = a \tan u$ to give

$$\sqrt{a^2 + x^2} = \sqrt{a^2 + a^2 \tan^2 u} = \sqrt{a^2 \sec^2 u} = |a \sec u|$$

- eliminate $\sqrt{x^2 - a^2}$ from an integrand by substituting $x = a \sec u$ to give

$$\sqrt{x^2 - a^2} = \sqrt{a^2 \sec^2 u - a^2} = \sqrt{a^2 \tan^2 u} = |a \tan u|$$

Be very careful with signs and absolute values when using this substitution. See Example 1.9.6.

When we have used substitutions before, we usually gave the new integration variable, u , as a function of the old integration variable x . Here we are doing the reverse — we are giving the old integration variable, x , in terms of the new integration variable u . We may do so, as long as we may invert to get u as a function of x . For example, with $x = a \sin u$, we may take $u = \arcsin \frac{x}{a}$. This is a good time for you to review the definitions of $\arcsin \theta$, $\arctan \theta$ and $\text{arcsec} \theta$. See Section 2.12, “Inverse Functions”, of the CLP-1 text.

As a warm-up, consider the area of a quarter of the unit circle.

✓ Example 1.9.1 Quarter of the unit circle

Compute the area of the unit circle lying in the first quadrant.

Solution

We know that the answer is $\frac{\pi}{4}$, but we can also compute this as an integral — we saw this way back in Example 1.1.16:

$$\text{area} = \int_0^1 \sqrt{1 - x^2} dx$$

- To simplify the integrand we substitute $x = \sin u$. With this choice $\frac{dx}{du} = \cos u$ and so $dx = \cos u du$.
- We also need to translate the limits of integration and it is perhaps easiest to do this by writing u as a function of x — namely $u(x) = \arcsin x$. Hence $u(0) = 0$ and $u(1) = \frac{\pi}{2}$.
- Hence the integral becomes

$$\begin{aligned} \int_0^1 \sqrt{1 - x^2} dx &= \int_0^{\frac{\pi}{2}} \sqrt{1 - \sin^2 u} \cdot \cos u du \\ &= \int_0^{\frac{\pi}{2}} \sqrt{\cos^2 u} \cdot \cos u du \\ &= \int_0^{\frac{\pi}{2}} \cos^2 u du \end{aligned}$$

Notice that here we have used that the *positive* square root $\sqrt{\cos^2 u} = |\cos u| = \cos u$ because $\cos(u) \geq 0$ for $0 \leq u \leq \frac{\pi}{2}$.

- To go further we use the techniques of Section 1.8.

$$\begin{aligned} \int_0^1 \sqrt{1-x^2} dx &= \int_0^{\frac{\pi}{2}} \cos^2 u du && \text{and since } \cos^2 u = \frac{1 + \cos 2u}{2} \\ &= \frac{1}{2} \int_0^{\frac{\pi}{2}} (1 + \cos(2u)) du \\ &= \frac{1}{2} \left[u + \frac{1}{2} \sin(2u) \right]_0^{\frac{\pi}{2}} \\ &= \frac{1}{2} \left(\frac{\pi}{2} - 0 + \frac{\sin \pi}{2} - \frac{\sin 0}{2} \right) \\ &= \frac{\pi}{4} \checkmark \end{aligned}$$

✓ Example 1.9.2 $\int \frac{x^2}{\sqrt{1-x^2}} dx$

Solution

We proceed much as we did in the previous example.

- To simplify the integrand we substitute $x = \sin u$. With this choice $\frac{dx}{du} = \cos u$ and so $dx = \cos u du$. Also note that $u = \arcsin x$.
- The integral becomes

$$\begin{aligned} \int \frac{x^2}{\sqrt{1-x^2}} dx &= \int \frac{\sin^2 u}{\sqrt{1-\sin^2 u}} \cdot \cos u du \\ &= \int \frac{\sin^2 u}{\sqrt{\cos^2 u}} \cdot \cos u du \end{aligned}$$

- To proceed further we need to get rid of the square-root. Since $u = \arcsin x$ has domain $-1 \leq x \leq 1$ and range $-\frac{\pi}{2} \leq u \leq \frac{\pi}{2}$, it follows that $\cos u \geq 0$ (since cosine is non-negative on these inputs). Hence

$$\sqrt{\cos^2 u} = \cos u \quad \text{when } -\frac{\pi}{2} \leq u \leq \frac{\pi}{2}$$

- So our integral now becomes

$$\begin{aligned} \int \frac{x^2}{\sqrt{1-x^2}} dx &= \int \frac{\sin^2 u}{\sqrt{\cos^2 u}} \cdot \cos u du \\ &= \int \frac{\sin^2 u}{\cos u} \cdot \cos u du \\ &= \int \sin^2 u du \\ &= \frac{1}{2} \int (1 - \cos 2u) du && \text{by Equation 1.8.4} \\ &= \frac{u}{2} - \frac{1}{4} \sin 2u + C \\ &= \frac{1}{2} \arcsin x - \frac{1}{4} \sin(2 \arcsin x) + C \end{aligned}$$

- We can simplify this further using a double-angle identity. Recall that $u = \arcsin x$ and that $x = \sin u$. Then

$$\sin 2u = 2 \sin u \cos u$$

We can replace $\cos u$ using $\cos^2 u = 1 - \sin^2 u$. Taking a square-root of this formula gives $\cos u = \pm\sqrt{1 - \sin^2 u}$. We need the positive branch here since $\cos u \geq 0$ when $-\frac{\pi}{2} \leq u \leq \frac{\pi}{2}$ (which is exactly the range of $\arcsin x$). Continuing along:

$$\begin{aligned}\sin 2u &= 2 \sin u \cdot \sqrt{1 - \sin^2 u} \\ &= 2x\sqrt{1 - x^2}\end{aligned}$$

Thus our solution is

$$\begin{aligned}\int \frac{x^2}{\sqrt{1 - x^2}} dx &= \frac{1}{2} \arcsin x - \frac{1}{4} \sin(2 \arcsin x) + C \\ &= \frac{1}{2} \arcsin x - \frac{1}{2} x \sqrt{1 - x^2} + C\end{aligned}$$

The above two examples illustrate the main steps of the approach. The next example is similar, but with more complicated limits of integration.

✓ Example 1.9.3 $\int_a^r \sqrt{r^2 - x^2} dx$

Let's find the area of the shaded region in the sketch below.



We'll set up the integral using vertical strips. The strip in the figure has width dx and height $\sqrt{r^2 - x^2}$. So the area is given by the integral

$$\text{area} = \int_a^r \sqrt{r^2 - x^2} dx$$

Which is very similar to the previous example.

Solution

- To evaluate the integral we substitute

$$x = x(u) = r \sin u \quad dx = \frac{dx}{du} du = r \cos u du$$

It is also helpful to write u as a function of x — namely $u = \arcsin \frac{x}{r}$.

- The integral runs from $x = a$ to $x = r$. These correspond to

$$\begin{aligned}u(r) &= \arcsin \frac{r}{r} = \arcsin 1 = \frac{\pi}{2} \\ u(a) &= \arcsin \frac{a}{r} \quad \text{which does not simplify further}\end{aligned}$$

- The integral then becomes

$$\begin{aligned}\int_a^r \sqrt{r^2 - x^2} dx &= \int_{\arcsin(a/r)}^{\pi/2} \sqrt{r^2 - r^2 \sin^2 u} \cdot r \cos u du \\ &= \int_{\arcsin(a/r)}^{\pi/2} r^2 \sqrt{1 - \sin^2 u} \cdot \cos u du \\ &= r^2 \int_{\arcsin(a/r)}^{\pi/2} \sqrt{\cos^2 u} \cdot \cos u du\end{aligned}$$

To proceed further (as we did in Examples 1.9.1 and 1.9.2) we need to think about whether $\cos u$ is positive or negative.

- Since a (as shown in the diagram) satisfies $0 \leq a \leq r$, we know that $u(a)$ lies between $\arcsin(0) = 0$ and $\arcsin(1) = \frac{\pi}{2}$. Hence the variable u lies between 0 and $\frac{\pi}{2}$, and on this range $\cos u \geq 0$. This allows us get rid of the square-root:

$$\sqrt{\cos^2 u} = |\cos u| = \cos u$$

- Putting this fact into our integral we get

$$\begin{aligned}\int_a^r \sqrt{r^2 - x^2} dx &= r^2 \int_{\arcsin(a/r)}^{\pi/2} \sqrt{\cos^2 u} \cdot \cos u du \\ &= r^2 \int_{\arcsin(a/r)}^{\pi/2} \cos^2 u du\end{aligned}$$

Recall the identity $\cos^2 u = \frac{1 + \cos 2u}{2}$ from Section 1.8

$$\begin{aligned}&= \frac{r^2}{2} \int_{\arcsin(a/r)}^{\pi/2} (1 + \cos 2u) du \\ &= \frac{r^2}{2} \left[u + \frac{1}{2} \sin(2u) \right]_{\arcsin(a/r)}^{\pi/2} \\ &= \frac{r^2}{2} \left(\frac{\pi}{2} + \frac{1}{2} \sin \pi - \arcsin(a/r) - \frac{1}{2} \sin(2 \arcsin(a/r)) \right) \\ &= \frac{r^2}{2} \left(\frac{\pi}{2} - \arcsin(a/r) - \frac{1}{2} \sin(2 \arcsin(a/r)) \right)\end{aligned}$$

Oof! But there is a little further to go before we are done.

- We can again simplify the term $\sin(2 \arcsin(a/r))$ using a double angle identity. Set $\theta = \arcsin(a/r)$. Then θ is the angle in the triangle on the right below. By the double angle formula for $\sin(2\theta)$ (Equation 1.8.2)

$$\begin{aligned}\sin(2\theta) &= 2 \sin \theta \cos \theta \\ &= 2 \frac{a}{r} \frac{\sqrt{r^2 - a^2}}{r}.\end{aligned}$$



- So finally the area is

$$\begin{aligned}\text{area} &= \int_a^r \sqrt{r^2 - x^2} dx \\ &= \frac{r^2}{2} \left(\frac{\pi}{2} - \arcsin(a/r) - \frac{1}{2} \sin(2 \arcsin(a/r)) \right) \\ &= \frac{\pi r^2}{4} - \frac{r^2}{2} \arcsin(a/r) - \frac{a}{2} \sqrt{r^2 - a^2}\end{aligned}$$

- This is a relatively complicated formula, but we can make some “reasonableness” checks, by looking at special values of a .
 - If $a = 0$ the shaded region, in the figure at the beginning of this example, is exactly one quarter of a disk of radius r and so has area $\frac{1}{4} \pi r^2$. Substituting $a = 0$ into our answer does indeed give $\frac{1}{4} \pi r^2$.
 - At the other extreme, if $a = r$, the shaded region disappears completely and so has area 0. Subbing $a = r$ into our answer does indeed give 0, since $\arcsin 1 = \frac{\pi}{2}$.

✓ Example 1.9.4 $\int_a^r x \sqrt{r^2 - x^2} dx$

The integral $\int_a^r x \sqrt{r^2 - x^2} dx$ looks a lot like the integral we just did in the previous 3 examples. It can also be evaluated using the trigonometric substitution $x = r \sin u$ — but that is unnecessarily complicated. Just because you have now learned how to use trigonometric substitution ¹ doesn't mean that you should forget everything you learned before.

Solution

This integral is *much* more easily evaluated using the simple substitution $u = r^2 - x^2$.

- Set $u = r^2 - x^2$. Then $du = -2x dx$, and so

$$\begin{aligned} \int_a^r x \sqrt{r^2 - x^2} dx &= \int_{r^2 - a^2}^0 \sqrt{u} \frac{du}{-2} \\ &= -\frac{1}{2} \left[\frac{u^{3/2}}{3/2} \right]_{r^2 - a^2}^0 \\ &= \frac{1}{3} [r^2 - a^2]^{3/2} \end{aligned}$$

Enough sines and cosines — let us try a tangent substitution.

✓ Example 1.9.5 $\int \frac{dx}{x^2 \sqrt{9+x^2}}$

Solution

As per our guidelines at the start of this section, the presence of the square root term $\sqrt{3^2+x^2}$ tells us to substitute $x = 3 \tan u$.

- Substitute

$$x = 3 \tan u \quad dx = 3 \sec^2 u \, du$$

This allows us to remove the square root:

$$\sqrt{9+x^2} = \sqrt{9+9 \tan^2 u} = 3\sqrt{1+\tan^2 u} = 3\sqrt{\sec^2 u} = 3|\sec u|$$

- Hence our integral becomes

$$\int \frac{dx}{x^2 \sqrt{9+x^2}} = \int \frac{3 \sec^2 u}{9 \tan^2 u \cdot 3 |\sec u|} du$$

- To remove the absolute value we must consider the range of values of u in the integral. Since $x = 3 \tan u$ we have $u = \arctan(x/3)$. The range² of arctangent is $-\frac{\pi}{2} \leq \arctan \leq \frac{\pi}{2}$ and so $u = \arctan(x/3)$ will always lie between $-\frac{\pi}{2}$ and $+\frac{\pi}{2}$. Hence $\cos u$ will always be positive, which in turn implies that $|\sec u| = \sec u$.
- Using this fact our integral becomes:

$$\begin{aligned} \int \frac{dx}{x^2 \sqrt{9+x^2}} &= \int \frac{3 \sec^2 u}{27 \tan^2 u |\sec u|} du \\ &= \frac{1}{9} \int \frac{\sec u}{\tan^2 u} du \quad \text{since } \sec u > 0 \end{aligned}$$

- Rewrite this in terms of sine and cosine

$$\begin{aligned} \int \frac{dx}{x^2 \sqrt{9+x^2}} &= \frac{1}{9} \int \frac{\sec u}{\tan^2 u} du \\ &= \frac{1}{9} \int \frac{1}{\cos u} \cdot \frac{\cos^2 u}{\sin^2 u} du = \frac{1}{9} \int \frac{\cos u}{\sin^2 u} du \end{aligned}$$

Now we can use the substitution rule with $y = \sin u$ and $dy = \cos u \, du$

$$\begin{aligned} &= \frac{1}{9} \int \frac{dy}{y^2} \\ &= -\frac{1}{9y} + C \\ &= -\frac{1}{9 \sin u} + C \end{aligned}$$

- The original integral was a function of x , so we still have to rewrite $\sin u$ in terms of x . Remember that $x = 3 \tan u$ or $u = \arctan(x/3)$. So u is the angle shown in the triangle below and we can read off the triangle that

$$\sin u = \frac{x}{\sqrt{9+x^2}}$$

$$\Rightarrow \int \frac{dx}{x^2\sqrt{9+x^2}} = -\frac{\sqrt{9+x^2}}{9x} + C$$

✓ Example 1.9.6 $\int \frac{x^2}{\sqrt{x^2-1}} dx$

Solution

This one requires a secant substitution, but otherwise is very similar to those above.

- Set $x = \sec u$ and $dx = \sec u \tan u du$. Then

$$\begin{aligned} \int \frac{x^2}{\sqrt{x^2-1}} dx &= \int \frac{\sec^2 u}{\sqrt{\sec^2 u - 1}} \sec u \tan u du \\ &= \int \sec^3 u \cdot \frac{\tan u}{\sqrt{\tan^2 u}} du \quad \text{since } \tan^2 u = \sec^2 u - 1 \\ &= \int \sec^3 u \cdot \frac{\tan u}{|\tan u|} du \end{aligned}$$

- As before we need to consider the range of u values in order to determine the sign of $\tan u$. Notice that the integrand is only defined when either $x < -1$ or $x > 1$; thus we should treat the cases $x < -1$ and $x > 1$ separately. Let us assume that $x > 1$ and we will come back to the case $x < -1$ at the end of the example.

When $x > 1$, our $u = \operatorname{arcsec} x$ takes values in $(0, \frac{\pi}{2})$. This follows since when $0 < u < \frac{\pi}{2}$, we have $0 < \cos u < 1$ and so $\sec u > 1$. Further, when $0 < u < \frac{\pi}{2}$, we have $\tan u > 0$. Thus $|\tan u| = \tan u$.

- Back to our integral, when $x > 1$:

$$\begin{aligned} \int \frac{x^2}{\sqrt{x^2-1}} dx &= \int \sec^3 u \cdot \frac{\tan u}{|\tan u|} du \\ &= \int \sec^3 u du \quad \text{since } \tan u \geq 0 \end{aligned}$$

This is exactly Example 1.8.22

$$= \frac{1}{2} \sec u \tan u + \frac{1}{2} \log |\sec u + \tan u| + C$$

- Since we started with a function of x we need to finish with one. We know that $\sec u = x$ and then we can use trig identities

$$\tan^2 u = \sec^2 u - 1 = x^2 - 1 \quad \text{so } \tan u = \pm \sqrt{x^2 - 1}$$

but we know

$$\tan u \geq 0 \quad \text{so } \tan u = \sqrt{x^2 - 1}$$

Thus

$$\int \frac{x^2}{\sqrt{x^2-1}} dx = \frac{1}{2} x \sqrt{x^2-1} + \frac{1}{2} \log |x + \sqrt{x^2-1}| + C$$

- The above holds when $x > 1$. We can confirm that it is also true when $x < -1$ by showing the right-hand side is a valid antiderivative of the integrand. To do so we must differentiate our answer. Notice that we do not need to consider the sign of $x + \sqrt{x^2-1}$ when we differentiate since we have already seen that

$$\frac{d}{dx} \log |x| = \frac{1}{x}$$

when either $x < 0$ or $x > 0$. So the following computation applies to both $x > 1$ and $x < -1$. The expressions become quite long so we differentiate each term separately:

$$\begin{aligned} \frac{d}{dx} \left[x\sqrt{x^2-1} \right] &= \left[\sqrt{x^2-1} + \frac{x^2}{\sqrt{x^2-1}} \right] \\ &= \frac{1}{\sqrt{x^2-1}} [(x^2-1) + x^2] \\ \frac{d}{dx} \log |x + \sqrt{x^2-1}| &= \frac{1}{x + \sqrt{x^2-1}} \cdot \left[1 + \frac{x}{\sqrt{x^2-1}} \right] \\ &= \frac{1}{x + \sqrt{x^2-1}} \cdot \frac{x + \sqrt{x^2-1}}{\sqrt{x^2-1}} \\ &= \frac{1}{\sqrt{x^2-1}} \end{aligned}$$

Putting things together then gives us

$$\begin{aligned} \frac{d}{dx} \left[\frac{1}{2}x\sqrt{x^2-1} + \frac{1}{2}\log |x + \sqrt{x^2-1}| + C \right] \\ &= \frac{1}{2\sqrt{x^2-1}} [(x^2-1) + x^2 + 1] + 0 \\ &= \frac{x^2}{\sqrt{x^2-1}} \end{aligned}$$

This tells us that our answer for $x > 1$ is also valid when $x < -1$ and so

$$\int \frac{x^2}{\sqrt{x^2-1}} dx = \frac{1}{2}x\sqrt{x^2-1} + \frac{1}{2}\log |x + \sqrt{x^2-1}| + C$$

when $x < -1$ and when $x > 1$.

In this example, we were lucky. The answer that we derived for $x > 1$ happened to be also valid for $x < -1$. This does not always happen with the $x = a \sec u$ substitution. When it doesn't, we have to apply separate $x > a$ and $x < -a$ analyses that are very similar to our $x > 1$ analysis above. Of course that doubles the tedium. So in the CLP-2 problem book, we will not pose questions that require separate $x > a$ and $x < -a$ computations.

The method, as we have demonstrated it above, works when our integrand contains the square root of very specific families of quadratic polynomials. In fact, the same method works for more general quadratic polynomials — all we need to do is complete the square³.

✓ Example 1.9.7 $\int_3^5 \frac{\sqrt{x^2-2x-3}}{x-1} dx$

This time we have an integral with a square root in the integrand, but the argument of the square root, while a quadratic function of x , is not in one of the standard forms $\sqrt{a^2-x^2}$, $\sqrt{a^2+x^2}$, $\sqrt{x^2-a^2}$. The reason that it is not in one of those forms is that the argument, x^2-2x-3 , contains a term, namely $-2x$ that is of degree one in x . So we try to manipulate it into one of the standard forms by completing the square.

Solution

- We first rewrite the quadratic polynomial x^2-2x-3 in the form $(x-a)^2+b$ for some constants a, b . The easiest way to do this is to expand both expressions and compare coefficients of x :

$$x^2-2x-3 = (x-a)^2+b = (x^2-2ax+a^2)+b$$

So — if we choose $-2a = -2$ (so the coefficients of x^1 match) and $a^2+b = -3$ (so the coefficients of x^0 match), then both expressions are equal. Hence we set $a = 1$ and $b = -4$. That is

$$x^2-2x-3 = (x-1)^2-4$$

Many of you may have seen this method when learning to sketch parabolas.

- Once this is done we can convert the square root of the integrand into a standard form by making the simple substitution $y = x - 1$. Here goes

$$\begin{aligned} & \int_3^5 \frac{\sqrt{x^2 - 2x - 3}}{x - 1} dx \\ &= \int_3^5 \frac{\sqrt{(x-1)^2 - 4}}{x - 1} dx \\ &= \int_2^4 \frac{\sqrt{y^2 - 4}}{y} dy && \text{with } y = x - 1, dy = dx \\ &= \int_0^{\pi/3} \frac{\sqrt{4 \sec^2 u - 4}}{2 \sec u} 2 \sec u \tan u du && \text{with } y = 2 \sec u \\ & && \text{and } dy = 2 \sec u \tan u du \end{aligned}$$

Notice that we could also do this in fewer steps by setting $(x - 1) = 2 \sec u$, $dx = 2 \sec u \tan u du$.

- To get the limits of integration we used that
 - the value of u that corresponds to $y = 2$ obeys $2 = y = 2 \sec u = \frac{2}{\cos u}$ or $\cos u = 1$, so that $u = 0$ works and
 - the value of u that corresponds to $y = 4$ obeys $4 = y = 2 \sec u = \frac{2}{\cos u}$ or $\cos u = \frac{1}{2}$, so that $u = \frac{\pi}{3}$ works.
- Now returning to the evaluation of the integral, we simplify and continue.

$$\begin{aligned} \int_3^5 \frac{\sqrt{x^2 - 2x - 3}}{x - 1} dx &= \int_0^{\pi/3} 2\sqrt{\sec^2 u - 1} \tan u du \\ &= 2 \int_0^{\pi/3} \tan^2 u du \quad \text{since } \sec^2 u = 1 + \tan^2 u \end{aligned}$$

In taking the square root of $\sec^2 u - 1 = \tan^2 u$ we used that $\tan u \geq 0$ on the range $0 \leq u \leq \frac{\pi}{3}$.

$$\begin{aligned} &= 2 \int_0^{\pi/3} [\sec^2 u - 1] du \quad \text{since } \sec^2 u = 1 + \tan^2 u, \text{ again} \\ &= 2 \left[\tan u - u \right]_0^{\pi/3} \\ &= 2 \left[\sqrt{3} - \frac{\pi}{3} \right] \end{aligned}$$

Exercises

Recall that we are using $\log x$ to denote the logarithm of x with base e . In other courses it is often denoted $\ln x$.

Stage 1

? 1 (*)

For each of the following integrals, choose the substitution that is most beneficial for evaluating the integral.

- $\int \frac{2x^2}{\sqrt{9x^2 - 16}} dx$
- $\int \frac{x^4 - 3}{\sqrt{1 - 4x^2}} dx$
- $\int (25 + x^2)^{-5/2} dx$

? 2

For each of the following integrals, choose a trigonometric substitution that will eliminate the roots.

1. $\int \frac{1}{\sqrt{x^2 - 4x + 1}} dx$
2. $\int \frac{(x-1)^6}{(-x^2 + 2x + 4)^{3/2}} dx$
3. $\int \frac{1}{\sqrt{4x^2 + 6x + 10}} dx$
4. $\int \sqrt{x^2 - x} dx$

? 3

In each part of this question, assume θ is an angle in the interval $[0, \pi/2]$.

1. If $\sin \theta = \frac{1}{20}$, what is $\cos \theta$?
2. If $\tan \theta = 7$, what is $\csc \theta$?
3. If $\sec \theta = \frac{\sqrt{x-1}}{2}$, what is $\tan \theta$?

? 4

Simplify the following expressions.

1. $\sin(\arccos(\frac{x}{2}))$
2. $\sin(\arctan(\frac{1}{\sqrt{3}}))$
3. $\sec(\arcsin(\sqrt{x}))$

Stage 2

? 5 (*)

Evaluate $\int \frac{1}{(x^2 + 4)^{3/2}} dx$.

? 6 (*)

Evaluate $\int_0^4 \frac{1}{(4 + x^2)^{3/2}} dx$. Your answer may not contain inverse trigonometric functions.

? 7 (*)

Evaluate $\int_0^{5/2} \frac{dx}{\sqrt{25 - x^2}}$.

? 8 (*)

Evaluate $\int \frac{dx}{\sqrt{x^2 + 25}}$. You may use that $\int \sec x dx = \log |\sec x + \tan x| + C$.

? 9

Evaluate $\int \frac{x+1}{\sqrt{2x^2 + 4x}} dx$.

? 10 (*)

Evaluate $\int \frac{dx}{x^2\sqrt{x^2+16}}$.

? 11 (*)

Evaluate $\int \frac{dx}{x^2\sqrt{x^2-9}}$ for $x \geq 3$. Do not include any inverse trigonometric functions in your answer.

? 12 (*)

(a) Show that $\int_0^{\pi/4} \cos^4 \theta d\theta = (8 + 3\pi)/32$.

(b) Evaluate $\int_{-1}^1 \frac{dx}{(x^2+1)^3}$.

? 13

Evaluate $\int_{-\pi/12}^{\pi/12} \frac{15x^3}{(x^2+1)(9-x^2)^{5/2}} dx$.

? 14 (*)

Evaluate $\int \sqrt{4-x^2} dx$.

? 15 (*)

Evaluate $\int \frac{\sqrt{25x^2-4}}{x} dx$ for $x > \frac{2}{5}$.

? 16

Evaluate $\int_{\sqrt{10}}^{\sqrt{17}} \frac{x^3}{\sqrt{x^2-1}} dx$.

? 17 (*)

Evaluate $\int \frac{dx}{\sqrt{3-2x-x^2}}$.

? 18

Evaluate $\int \frac{1}{(2x-3)^3\sqrt{4x^2-12x+8}} dx$ for $x > 2$.

? 19

Evaluate $\int_0^1 \frac{x^2}{(x^2+1)^{3/2}} dx$.

You may use that $\int \sec x dx = \log |\sec x + \tan x| + C$.

? 20

Evaluate $\int \frac{1}{(x^2 + 1)^2} dx$.

Stage 3

? 21

Evaluate $\int \frac{x^2}{\sqrt{x^2 - 2x + 2}} dx$.

You may assume without proof that $\int \sec^3 \theta d\theta = \frac{1}{2} \sec \theta \tan \theta + \frac{1}{2} \log |\sec \theta + \tan \theta| + C$.

? 22

Evaluate $\int \frac{1}{\sqrt{3x^2 + 5x}} dx$.

You may use that $\int \sec x dx = \log |\sec x + \tan x| + C$.

? 23

Evaluate $\int \frac{(1+x^2)^{3/2}}{x} dx$. You may use the fact that $\int \csc \theta d\theta = \log |\cot \theta - \csc \theta| + C$.

? 24

Below is the graph of the ellipse $(\frac{x}{4})^2 + (\frac{y}{2})^2 = 1$. Find the area of the shaded region using the ideas from this section.



? 25

Let $f(x) = \frac{|x|}{\sqrt{1-x^2}}$, and let R be the region between $f(x)$ and the x -axis over the interval $[-\frac{1}{2}, \frac{1}{2}]$.

1. Find the area of R .
2. Find the volume of the solid formed by rotating R about the x -axis.

? 26

Evaluate $\int \sqrt{1+e^x} dx$. You may use the antiderivative $\int \csc \theta d\theta = \log |\cot \theta - \csc \theta| + C$.

? 27

Consider the following work.

$$\begin{aligned}
 \int \frac{1}{1-x^2} dx &= \int \frac{1}{1-\sin^2 \theta} \cos \theta d\theta && \text{using } x = \sin \theta, dx = \cos \theta d\theta \\
 &= \int \frac{\cos \theta}{\cos^2 \theta} d\theta \\
 &= \int \sec \theta d\theta \\
 &= \log |\sec \theta + \tan \theta| + C && \text{Example 1.8.19} \\
 &= \log \left| \frac{1}{\sqrt{1-x^2}} + \frac{x}{\sqrt{1-x^2}} \right| + C && \text{using the triangle below} \\
 &= \log \left| \frac{1+x}{\sqrt{1-x^2}} \right| + C
 \end{aligned}$$



1. Differentiate $\log \left| \frac{1+x}{\sqrt{1-x^2}} \right|$.
2. True or false: $\int_2^3 \frac{1}{1-x^2} dx = \left[\log \left| \frac{1+x}{\sqrt{1-x^2}} \right| \right]_{x=2}^{x=3}$
3. Was the work in the question correct? Explain.

? 28

1. Suppose we are evaluating an integral that contains the term $\sqrt{a^2 - x^2}$, where a is a positive constant, and we use the substitution $x = a \sin u$ (with inverse $u = \arcsin(x/a)$), so that

$$\sqrt{a^2 - x^2} = \sqrt{a^2 \cos^2 u} = |a \cos u|$$

Under what circumstances is $|a \cos u| \neq a \cos u$?

2. Suppose we are evaluating an integral that contains the term $\sqrt{a^2 + x^2}$, where a is a positive constant, and we use the substitution $x = a \tan u$ (with inverse $u = \arctan(x/a)$), so that

$$\sqrt{a^2 + x^2} = \sqrt{a^2 \sec^2 u} = |a \sec u|$$

Under what circumstances is $|a \sec u| \neq a \sec u$?

3. Suppose we are evaluating an integral that contains the term $\sqrt{x^2 - a^2}$, where a is a positive constant, and we use the substitution $x = a \sec u$ (with inverse $u = \operatorname{arcsec}(x/a) = \arccos(a/x)$), so that

$$\sqrt{x^2 - a^2} = \sqrt{a^2 \tan^2 u} = |a \tan u|$$

Under what circumstances is $|a \tan u| \neq a \tan u$?

1. To paraphrase the Law of the Instrument, possibly Mark Twain and definitely some psychologists, when you have a shiny new hammer, everything looks like a nail.
2. To be pedantic, we mean the range of the “standard” arctangent function or its “principle value”. One can define other arctangent functions with different ranges.
3. If you have not heard of “completing the square” don't worry. It is not a difficult method and it will only take you a few moments to learn. It refers to rewriting a quadratic polynomial $P(x) = ax^2 + bx + c$ as $P(x) = a(x+d)^2 + e$ for new constants d, e .

1.10: Partial Fractions

Partial fractions is the name given to a technique of integration that may be used to integrate any rational function¹. We already know how to integrate some simple rational functions

$$\int \frac{1}{x} dx = \log|x| + C \quad \int \frac{1}{1+x^2} dx = \arctan(x) + C$$

Combining these with the substitution rule, we can integrate similar but more complicated rational functions:

$$\int \frac{1}{2x+3} dx = \frac{1}{2} \log|2x+3| + C \quad \int \frac{1}{3+4x^2} dx = \frac{1}{2\sqrt{3}} \arctan\left(\frac{2x}{\sqrt{3}}\right) + C$$

By summing such terms together we can integrate yet more complicated forms

$$\int \left[x + \frac{1}{x+1} + \frac{1}{x-1} \right] dx = \frac{x^2}{2} + \log|x+1| + \log|x-1| + C$$

However we are not (typically) presented with a rational function nicely decomposed into neat little pieces. It is far more likely that the rational function will be written as the ratio of two polynomials. For example:

$$\int \frac{x^3+x}{x^2-1} dx$$

In this specific example it is not hard to confirm that

$$x + \frac{1}{x+1} + \frac{1}{x-1} = \frac{x(x+1)(x-1) + (x-1) + (x+1)}{(x+1)(x-1)} = \frac{x^3+x}{x^2-1}$$

and hence

$$\begin{aligned} \int \frac{x^3+x}{x^2-1} dx &= \int \left[x + \frac{1}{x+1} + \frac{1}{x-1} \right] dx \\ &= \frac{x^2}{2} + \log|x+1| + \log|x-1| + C \end{aligned}$$

Of course going in this direction (from a sum of terms to a single rational function) is straightforward. To be useful we need to understand how to do this in reverse: decompose a given rational function into a sum of simpler pieces that we can integrate.

Suppose that $N(x)$ and $D(x)$ are polynomials. The basic strategy is to write $\frac{N(x)}{D(x)}$ as a sum of very simple, easy to integrate rational functions, namely

1. polynomials — we shall see below that these are needed when the degree² of $N(x)$ is equal to or strictly bigger than the degree of $D(x)$, and
2. rational functions of the particularly simple form $\frac{A}{(ax+b)^n}$ and
3. rational functions of the form $\frac{Ax+B}{(ax^2+bx+c)^m}$.

We already know how to integrate the first two forms, and we'll see how to integrate the third form in the near future.

To begin to explore this method of decomposition, let us go back to the example we just saw

$$x + \frac{1}{x+1} + \frac{1}{x-1} = \frac{x(x+1)(x-1) + (x-1) + (x+1)}{(x+1)(x-1)} = \frac{x^3+x}{x^2-1}$$

The technique that we will use is based on two observations:

1. The denominators on the left-hand side of are the factors of the denominator $x^2-1 = (x-1)(x+1)$ on the right-hand side.
2. Use $P(x)$ to denote the polynomial on the left hand side, and then use $N(x)$ and $D(x)$ to denote the numerator and denominator of the right hand side. That is

$$P(x) = x \quad N(x) = x^3+x \quad D(x) = x^2-1.$$

Then the degree of $N(x)$ is the sum of the degrees of $P(x)$ and $D(x)$. This is because the highest degree term in $N(x)$ is x^3 , which comes from multiplying $P(x)$ by $D(x)$, as we see in

$$x + \frac{1}{x+1} + \frac{1}{x-1} = \frac{\overbrace{x}^{P(x)} \overbrace{(x+1)(x-1)}^{D(x)} + (x-1) + (x+1)}{(x+1)(x-1)} = \frac{x^3 + x}{x^2 - 1}$$

More generally, the presence of a polynomial on the left hand side is signalled on the right hand side by the fact that the degree of the numerator is at least as large as the degree of the denominator.

Partial fraction decomposition examples

Rather than writing up the technique — known as the partial fraction decomposition — in full generality, we will instead illustrate it through a sequence of examples.

✓ Example 1.10.1 $\int \frac{x-3}{x^2-3x+2} dx$

In this example, we integrate $\frac{N(x)}{D(x)} = \frac{x-3}{x^2-3x+2}$.

Solution

- *Step 1.* We first check to see if a polynomial $P(x)$ is needed. To do so, we check to see if the degree of the numerator, $N(x)$, is strictly smaller than the degree of the denominator $D(x)$. In this example, the numerator, $x - 3$, has degree one and that is indeed strictly smaller than the degree of the denominator, $x^2 - 3x + 2$, which is two. In this case ³ we do not need to extract a polynomial $P(x)$ and we move on to step 2.
- *Step 2.* The second step is to factor the denominator

$$x^2 - 3x + 2 = (x - 1)(x - 2)$$

In this example it is quite easy, but in future examples (and quite possibly in your homework, quizzes and exam) you will have to work harder to factor the denominator. In Appendix A.16 we have written up some simple tricks for factoring polynomials. We will illustrate them in Example 1.10.3 below.

- *Step 3.* The third step is to write $\frac{x-3}{x^2-3x+2}$ in the form

$$\frac{x-3}{x^2-3x+2} = \frac{A}{x-1} + \frac{B}{x-2}$$

for some constants A and B . More generally, if the denominator consists of n different linear factors, then we decompose the ratio as

$$\text{rational function} = \frac{A_1}{\text{linear factor 1}} + \frac{A_2}{\text{linear factor 2}} + \dots + \frac{A_n}{\text{linear factor n}}$$

To proceed we need to determine the values of the constants A , B and there are several different methods to do so. Here are two methods

- *Step 3 — Algebra Method.* This approach has the benefit of being conceptually clearer and easier, but the downside is that it is more tedious.

To determine the values of the constants A , B , we put⁴ the right-hand side back over the common denominator $(x - 1)(x - 2)$.

$$\frac{x-3}{x^2-3x+2} = \frac{A}{x-1} + \frac{B}{x-2} = \frac{A(x-2) + B(x-1)}{(x-1)(x-2)}$$

The fraction on the far left is the same as the fraction on the far right if and only if their numerators are the same.

$$x - 3 = A(x - 2) + B(x - 1)$$

Write the right hand side as a polynomial in standard form (i.e. collect up all x terms and all constant terms)

$$x - 3 = (A + B)x + (-2A - B)$$

For these two polynomials to be the same, the coefficient of x on the left hand side and the coefficient of x on the right hand side must be the same. Similarly the coefficients of x^0 (i.e. the constant terms) must match. This gives us a system of two equations.

$$A + B = 1 \quad -2A - B = -3$$

in the two unknowns A, B . We can solve this system by

- using the first equation, namely $A + B = 1$, to determine A in terms of B :

$$A = 1 - B$$

- Substituting this into the remaining equation eliminates the A from second equation, leaving one equation in the one unknown B , which can then be solved for B :

$$\begin{array}{ll} -2A - B = -3 & \text{substitute } A = 1 - B \\ -2(1 - B) - B = -3 & \text{clean up} \\ -2 + B = -3 & \text{so } B = -1 \end{array}$$

- Once we know B , we can substitute it back into $A = 1 - B$ to get A .

$$A = 1 - B = 1 - (-1) = 2$$

Hence

$$\frac{x - 3}{x^2 - 3x + 2} = \frac{2}{x - 1} - \frac{1}{x - 2}$$

- Step 3 — Sneaky Method.* This takes a little more work to understand, but it is more efficient than the algebra method.

We wish to find A and B for which

$$\frac{x - 3}{(x - 1)(x - 2)} = \frac{A}{x - 1} + \frac{B}{x - 2}$$

Note that the denominator on the left hand side has been written in factored form.

- To determine A , we multiply both sides of the equation by A 's denominator, which is $x - 1$,

$$\frac{x - 3}{x - 2} = A + \frac{(x - 1)B}{x - 2}$$

and then we completely eliminate B from the equation by evaluating at $x = 1$. This value of x is chosen to make $x - 1 = 0$.

$$\left. \frac{x - 3}{x - 2} \right|_{x=1} = A + \left. \frac{(x - 1)B}{x - 2} \right|_{x=1} \implies A = \frac{1 - 3}{1 - 2} = 2$$

- To determine B , we multiply both sides of the equation by B 's denominator, which is $x - 2$,

$$\frac{x - 3}{x - 1} = \frac{(x - 2)A}{x - 1} + B$$

and then we completely eliminate A from the equation by evaluating at $x = 2$. This value of x is chosen to make $x - 2 = 0$.

$$\left. \frac{x - 3}{x - 1} \right|_{x=2} = \left. \frac{(x - 2)A}{x - 1} \right|_{x=2} + B \implies B = \frac{2 - 3}{2 - 1} = -1$$

Hence we have (the thankfully consistent answer)

$$\frac{x - 3}{x^2 - 3x + 2} = \frac{2}{x - 1} - \frac{1}{x - 2}$$

Notice that no matter which method we use to find the constants we can easily check our answer by summing the terms back together:

$$\begin{aligned}\frac{2}{x-1} - \frac{1}{x-2} &= \frac{2(x-2) - (x-1)}{(x-2)(x-1)} \\ &= \frac{2x-4-x+1}{x^2-3x+2} = \frac{x-3}{x^2-3x+2} \checkmark\end{aligned}$$

- *Step 4.* The final step is to integrate.

$$\begin{aligned}\int \frac{x-3}{x^2-3x+2} dx &= \int \frac{2}{x-1} dx + \int \frac{-1}{x-2} dx \\ &= 2 \log|x-1| - \log|x-2| + C\end{aligned}$$

Perhaps the first thing that you notice is that this process takes quite a few steps⁵. However no single step is all that complicated; it only takes practice. With that said, let's do another, slightly more complicated, one.

✓ Example 1.10.2 $\int \frac{3x^3-8x^2+4x-1}{x^2-3x+2} dx$

In this example, we integrate $\frac{N(x)}{D(x)} = \frac{3x^3-8x^2+4x-1}{x^2-3x+2}$.

Solution

- *Step 1.* We first check to see if the degree of the numerator $N(x)$ is strictly smaller than the degree of the denominator $D(x)$. In this example, the numerator, $3x^3 - 8x^2 + 4x - 1$, has degree three and the denominator, $x^2 - 3x + 2$, has degree two. As $3 \geq 2$, we have to implement the first step.

The goal of the first step is to write $\frac{N(x)}{D(x)}$ in the form

$$\frac{N(x)}{D(x)} = P(x) + \frac{R(x)}{D(x)}$$

with $P(x)$ being a polynomial and $R(x)$ being a polynomial of degree strictly smaller than the degree of $D(x)$. The right hand side is $\frac{P(x)D(x)+R(x)}{D(x)}$, so we have to express the numerator in the form $N(x) = P(x)D(x) + R(x)$, with $P(x)$ and $R(x)$ being polynomials and with the degree of R being strictly smaller than the degree of D . $P(x)D(x)$ is a sum of expressions of the form $ax^n D(x)$. We want to pull as many expressions of this form as possible out of the numerator $N(x)$, leaving only a low degree remainder $R(x)$.

We do this using long division — the same long division you learned in school, but with the base 10 replaced by x .

- We start by observing that to get from the highest degree term in the denominator (x^2) to the highest degree term in the numerator ($3x^3$), we have to multiply it by $3x$. So we write,



In the above expression, the denominator is on the left, the numerator is on the right and $3x$ is written above the highest order term of the numerator. Always put lower powers of x to the right of higher powers of x — this mirrors how you do long division with numbers; lower powers of ten sit to the right of higher powers of ten.

- Now we subtract $3x$ times the denominator, $x^2 - 3x + 2$, which is $3x^3 - 9x^2 + 6x$, from the numerator.



- This has left a remainder of $x^2 - 2x - 1$. To get from the highest degree term in the denominator (x^2) to the highest degree term in the remainder (x^2), we have to multiply by 1. So we write,



- Now we subtract 1 times the denominator, $x^2 - 3x + 2$, which is $x^2 - 3x + 2$, from the remainder.



- This leaves a remainder of $x - 3$. Because the remainder has degree 1, which is smaller than the degree of the denominator (being degree 2), we stop.

- In this example, when we subtracted $3x(x^2 - 3x + 2)$ and $1(x^2 - 3x + 2)$ from $3x^3 - 8x^2 + 4x - 1$ we ended up with $x - 3$. That is,

$$\begin{aligned}3x^3 - 8x^2 + 4x - 1 - 3x(x^2 - 3x + 2) - 1(x^2 - 3x + 2) \\ = x - 3\end{aligned}$$

or, collecting the two terms proportional to $(x^2 - 3x + 2)$


$$3x^3 - 8x^2 + 4x - 1 - (3x + 1)(x^2 - 3x + 2) = x - 3$$

Moving the $(3x + 1)(x^2 - 3x + 2)$ to the right hand side and dividing the whole equation by $x^2 - 3x + 2$ gives

$$\frac{3x^3 - 8x^2 + 4x - 1}{x^2 - 3x + 2} = 3x + 1 + \frac{x - 3}{x^2 - 3x + 2}$$

And we can easily check this expression just by summing the two terms on the right-hand side.

We have written the integrand in the form $\frac{N(x)}{D(x)} = P(x) + \frac{R(x)}{D(x)}$, with the degree of $R(x)$ strictly smaller than the degree of $D(x)$, which is what we wanted. Observe that $R(x)$ is the final remainder of the long division procedure and $P(x)$ is at the top of the long division computation. This is the end of Step 1. Oof! You should definitely practice this step.

- 
- Step 2. The second step is to factor the denominator

$$x^2 - 3x + 2 = (x - 1)(x - 2)$$

We already did this in Example 1.10.1.

- Step 3. The third step is to write $\frac{x-3}{x^2-3x+2}$ in the form

$$\frac{x - 3}{x^2 - 3x + 2} = \frac{A}{x - 1} + \frac{B}{x - 2}$$

for some constants A and B . We already did this in Example 1.10.1. We found $A = 2$ and $B = -1$.

- Step 4. The final step is to integrate.

$$\begin{aligned} \int \frac{3x^3 - 8x^2 + 4x - 1}{x^2 - 3x + 2} dx &= \int [3x + 1] dx + \int \frac{2}{x - 1} dx + \int \frac{-1}{x - 2} dx \\ &= \frac{3}{2}x^2 + x + 2 \log|x - 1| - \log|x - 2| + C \end{aligned}$$

You can see that the integration step is quite quick — almost all the work is in preparing the integrand.

Here is a very solid example. It is quite long and the steps are involved. However please persist. No single step is too difficult.

✓ Example 1.10.3 $\int \frac{x^4 + 5x^3 + 16x^2 + 26x + 22}{x^3 + 3x^2 + 7x + 5} dx$

In this example, we integrate $\frac{N(x)}{D(x)} = \frac{x^4 + 5x^3 + 16x^2 + 26x + 22}{x^3 + 3x^2 + 7x + 5}$.

Solution

- Step 1. Again, we start by comparing the degrees of the numerator and denominator. In this example, the numerator, $x^4 + 5x^3 + 16x^2 + 26x + 22$, has degree four and the denominator, $x^3 + 3x^2 + 7x + 5$, has degree three. As $4 \geq 3$, we must execute the first step, which is to write $\frac{N(x)}{D(x)}$ in the form

$$\frac{N(x)}{D(x)} = P(x) + \frac{R(x)}{D(x)}$$

with $P(x)$ being a polynomial and $R(x)$ being a polynomial of degree strictly smaller than the degree of $D(x)$. This step is accomplished by long division, just as we did in Example 1.10.3. We'll go through the whole process in detail again.

Actually — before you read on ahead, please have a go at the long division. It is good practice.

- We start by observing that to get from the highest degree term in the denominator (x^3) to the highest degree term in the numerator (x^4), we have to multiply by x . So we write,



- Now we subtract x times the denominator $x^3 + 3x^2 + 7x + 5$, which is $x^4 + 3x^3 + 7x^2 + 5x$, from the numerator.
- The remainder was $2x^3 + 9x^2 + 21x + 22$. To get from the highest degree term in the denominator (x^3) to the highest degree term in the remainder ($2x^3$), we have to multiply by 2. So we write,
- Now we subtract 2 times the denominator $x^3 + 3x^2 + 7x + 5$, which is $2x^3 + 6x^2 + 14x + 10$, from the remainder.
- This leaves a remainder of $3x^2 + 7x + 12$. Because the remainder has degree 2, which is smaller than the degree of the denominator, which is 3, we stop.
- In this example, when we subtracted $x(x^3 + 3x^2 + 7x + 5)$ and $2(x^3 + 3x^2 + 7x + 5)$ from $x^4 + 5x^3 + 16x^2 + 26x + 22$ we ended up with $3x^2 + 7x + 12$. That is,

$$\begin{aligned} x^4 + 5x^3 + 16x^2 + 26x + 22 &- x(x^3 + 3x^2 + 7x + 5) \\ &- 2(x^3 + 3x^2 + 7x + 5) \\ &= 3x^2 + 7x + 12 \end{aligned}$$

or, collecting the two terms proportional to $(x^3 + 3x^2 + 7x + 5)$ we get

$$\begin{aligned} x^4 + 5x^3 + 16x^2 + 26x + 22 &- (x + 2)(x^3 + 3x^2 + 7x + 5) \\ &= 3x^2 + 7x + 12 \end{aligned}$$

Moving the $(x + 2)(x^3 + 3x^2 + 7x + 5)$ to the right hand side and dividing the whole equation by $x^3 + 3x^2 + 7x + 5$ gives

$$\frac{x^4 + 5x^3 + 16x^2 + 26x + 22}{x^3 + 3x^2 + 7x + 5} = x + 2 + \frac{3x^2 + 7x + 12}{x^3 + 3x^2 + 7x + 5}$$

This is of the form $\frac{N(x)}{D(x)} = P(x) + \frac{R(x)}{D(x)}$, with the degree of $R(x)$ strictly smaller than the degree of $D(x)$, which is what we wanted. Observe, once again, that $R(x)$ is the final remainder of the long division procedure and $P(x)$ is at the top of the long division computation.

- Step 2.** The second step is to factor the denominator $D(x) = x^3 + 3x^2 + 7x + 5$. In the “real world” factorisation of polynomials is often very hard. Fortunately⁶, this is not the “real world” and there is a trick available to help us find this factorisation. The reader should take some time to look at Appendix A.16 before proceeding.
 - The trick exploits the fact that most polynomials that appear in homework assignments and on tests have integer coefficients and some integer roots. Any integer root of a polynomial that has integer coefficients, like $D(x) = x^3 + 3x^2 + 7x + 5$, must divide the constant term of the polynomial exactly. Why this is true is explained⁷ in Appendix A.16.
 - So any integer root of $x^3 + 3x^2 + 7x + 5$ must divide 5 exactly. Thus the only integers which can be roots of $D(x)$ are ± 1 and ± 5 . Of course, not all of these give roots of the polynomial — in fact there is no guarantee that any of them will be. We have to test each one.
 - To test if $+1$ is a root, we sub $x = 1$ into $D(x)$:

$$D(1) = 1^3 + 3(1)^2 + 7(1) + 5 = 16$$

As $D(1) \neq 0$, 1 is not a root of $D(x)$.

- To test if -1 is a root, we sub it into $D(x)$:

$$D(-1) = (-1)^3 + 3(-1)^2 + 7(-1) + 5 = -1 + 3 - 7 + 5 = 0$$

As $D(-1) = 0$, -1 is a root of $D(x)$. As -1 is a root of $D(x)$, $(x - (-1)) = (x + 1)$ must factor $D(x)$ exactly. We can factor the $(x + 1)$ out of $D(x) = x^3 + 3x^2 + 7x + 5$ by long division once again.

- Dividing $D(x)$ by $(x + 1)$ gives:

This time, when we subtracted $x^2(x + 1)$ and $2x(x + 1)$ and $5(x + 1)$ from $x^3 + 3x^2 + 7x + 5$ we ended up with 0 — as we knew would happen, because we knew that $x + 1$ divides $x^3 + 3x^2 + 7x + 5$ exactly. Hence

$$x^3 + 3x^2 + 7x + 5 - x^2(x+1) - 2x(x+1) - 5(x+1) = 0$$

or

$$x^3 + 3x^2 + 7x + 5 = x^2(x+1) + 2x(x+1) + 5(x+1)$$

or

$$x^3 + 3x^2 + 7x + 5 = (x^2 + 2x + 5)(x+1)$$

- It isn't quite time to stop yet; we should attempt to factor the quadratic factor, $x^2 + 2x + 5$. We can use the quadratic formula⁸ to find the roots of $x^2 + 2x + 5$:

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-2 \pm \sqrt{4 - 20}}{2} = \frac{-2 \pm \sqrt{-16}}{2}$$

Since this expression contains the square root of a negative number the equation $x^2 + 2x + 5 = 0$ has no real solutions; without the use of complex numbers, $x^2 + 2x + 5$ cannot be factored.

We have reached the end of step 2. At this point we have

$$\frac{x^4 + 5x^3 + 16x^2 + 26x + 22}{x^3 + 3x^2 + 7x + 5} = x + 2 + \frac{3x^2 + 7x + 12}{(x+1)(x^2 + 2x + 5)}$$

- Step 3.** The third step is to write $\frac{3x^2 + 7x + 12}{(x+1)(x^2 + 2x + 5)}$ in the form

$$\frac{3x^2 + 7x + 12}{(x+1)(x^2 + 2x + 5)} = \frac{A}{x+1} + \frac{Bx + C}{x^2 + 2x + 5}$$

for some constants A , B and C .

Note that the numerator, $Bx + C$ of the second term on the right hand side is not just a constant. It is of degree one, which is exactly one smaller than the degree of the denominator, $x^2 + 2x + 5$. More generally, if the denominator consists of n different linear factors and m different quadratic factors, then we decompose the ratio as

$$\begin{aligned} \text{rational function} &= \frac{A_1}{\text{linear factor 1}} + \frac{A_2}{\text{linear factor 2}} + \cdots + \frac{A_n}{\text{linear factor n}} \\ &+ \frac{B_1x + C_1}{\text{quadratic factor 1}} + \frac{B_2x + C_2}{\text{quadratic factor 2}} + \cdots \\ &+ \frac{B_mx + C_m}{\text{quadratic factor m}} \end{aligned}$$

To determine the values of the constants A , B , C , we put the right hand side back over the common denominator $(x+1)(x^2 + 2x + 5)$.

$$\begin{aligned} \frac{3x^2 + 7x + 12}{(x+1)(x^2 + 2x + 5)} &= \frac{A}{x+1} + \frac{Bx + C}{x^2 + 2x + 5} \\ &= \frac{A(x^2 + 2x + 5) + (Bx + C)(x+1)}{(x+1)(x^2 + 2x + 5)} \end{aligned}$$

The fraction on the far left is the same as the fraction on the far right if and only if their numerators are the same.

$$3x^2 + 7x + 12 = A(x^2 + 2x + 5) + (Bx + C)(x+1)$$

Again, as in Example 1.10.1, there are a couple of different ways to determine the values of A , B and C from this equation.

- Step 3 — Algebra Method.** The conceptually clearest procedure is to write the right hand side as a polynomial in standard form (i.e. collect up all x^2 terms, all x terms and all constant terms)

$$3x^2 + 7x + 12 = (A + B)x^2 + (2A + B + C)x + (5A + C)$$

For these two polynomials to be the same, the coefficient of x^2 on the left hand side and the coefficient of x^2 on the right hand side must be the same. Similarly the coefficients of x^1 must match and the coefficients of x^0 must match.

This gives us a system of three equations

$$A + B = 3 \quad 2A + B + C = 7 \quad 5A + C = 12$$

in the three unknowns A, B, C . We can solve this system by

- using the first equation, namely $A + B = 3$, to determine A in terms of B : $A = 3 - B$.
- Substituting $A = 3 - B$ into the remaining two equations eliminates the A 's from these two equations, leaving two equations in the two unknowns B and C .

$$\begin{aligned} A &= 3 - B & 2A + B + C &= 7 & 5A + C &= 12 \\ \Rightarrow & & 2(3 - B) + B + C &= 7 & 5(3 - B) + C &= 12 \\ \Rightarrow & & -B + C &= 1 & -5B + C &= -3 \end{aligned}$$

- Now we can use the equation $-B + C = 1$, to determine B in terms of C : $B = C - 1$.
- Substituting this into the remaining equation eliminates the B 's leaving an equation in the one unknown C , which is easy to solve.

$$\begin{aligned} B &= C - 1 & -5B + C &= -3 \\ \Rightarrow & & -5(C - 1) + C &= -3 \\ \Rightarrow & & -4C &= -8 \end{aligned}$$

- So $C = 2$, and then $B = C - 1 = 1$, and then $A = 3 - B = 2$. Hence

$$\frac{3x^2 + 7x + 12}{(x + 1)(x^2 + 2x + 5)} = \frac{2}{x + 1} + \frac{x + 2}{x^2 + 2x + 5}$$

- Step 3 — Sneaky Method.* While the above method is transparent, it is rather tedious. It is arguably better to use the second, sneakier and more efficient, procedure. In order for

$$3x^2 + 7x + 12 = A(x^2 + 2x + 5) + (Bx + C)(x + 1)$$

the equation must hold for all values of x .

- In particular, it must be true for $x = -1$. When $x = -1$, the factor $(x + 1)$ multiplying $Bx + C$ is exactly zero. So B and C disappear from the equation, leaving us with an easy equation to solve for A :

$$\begin{aligned} 3x^2 + 7x + 12 \Big|_{x=-1} &= \left[A(x^2 + 2x + 5) + (Bx + C)(x + 1) \right]_{x=-1} \\ \implies 8 = 4A &\implies A = 2 \end{aligned}$$

- Sub this value of A back in and simplify.

$$\begin{aligned} 3x^2 + 7x + 12 &= 2(x^2 + 2x + 5) + (Bx + C)(x + 1) \\ x^2 + 3x + 2 &= (Bx + C)(x + 1) \end{aligned}$$

Since $(x + 1)$ is a factor on the right hand side, it must also be a factor on the left hand side.

$$\begin{aligned} (x + 2)(x + 1) &= (Bx + C)(x + 1) \\ \Rightarrow (x + 2) &= (Bx + C) \Rightarrow B = 1, C = 2 \end{aligned}$$

So again we find that

$$\frac{3x^2 + 7x + 12}{(x + 1)(x^2 + 2x + 5)} = \frac{2}{x + 1} + \frac{x + 2}{x^2 + 2x + 5} \checkmark$$

Thus our integrand can be written as

$$\frac{x^4 + 5x^3 + 16x^2 + 26x + 22}{x^3 + 3x^2 + 7x + 5} = x + 2 + \frac{2}{x + 1} + \frac{x + 2}{x^2 + 2x + 5}$$

- *Step 4.* Now we can finally integrate! The first two pieces are easy.

$$\int (x+2) dx = \frac{1}{2}x^2 + 2x \quad \int \frac{2}{x+1} dx = 2 \log|x+1|$$

(We're leaving the arbitrary constant to the end of the computation.)

The final piece is a little harder. The idea is to complete the square⁹ in the denominator

$$\frac{x+2}{x^2+2x+5} = \frac{x+2}{(x+1)^2+4}$$

and then make a change of variables to make the fraction look like $\frac{ay+b}{y^2+1}$. In this case

$$\frac{x+2}{(x+1)^2+4} = \frac{1}{4} \frac{x+2}{\left(\frac{x+1}{2}\right)^2+1}$$

so we make the change of variables $y = \frac{x+1}{2}$, $dy = \frac{dx}{2}$, $x = 2y - 1$, $dx = 2 dy$

$$\begin{aligned} \int \frac{x+2}{(x+1)^2+4} dx &= \frac{1}{4} \int \frac{x+2}{\left(\frac{x+1}{2}\right)^2+1} dx \\ &= \frac{1}{4} \int \frac{(2y-1)+2}{y^2+1} 2 dy = \frac{1}{2} \int \frac{2y+1}{y^2+1} dy \\ &= \int \frac{y}{y^2+1} dy + \frac{1}{2} \int \frac{1}{y^2+1} dy \end{aligned}$$

Both integrals are easily evaluated, using the substitution $u = y^2 + 1$, $du = 2y dy$ for the first.

$$\begin{aligned} \int \frac{y}{y^2+1} dy &= \int \frac{1}{u} \frac{du}{2} = \frac{1}{2} \log|u| = \frac{1}{2} \log(y^2+1) \\ &= \frac{1}{2} \log\left[\left(\frac{x+1}{2}\right)^2+1\right] \\ \frac{1}{2} \int \frac{1}{y^2+1} dy &= \frac{1}{2} \arctan y = \frac{1}{2} \arctan\left(\frac{x+1}{2}\right) \end{aligned}$$

That's finally it. Putting all of the pieces together

$$\begin{aligned} \int \frac{x^4+5x^3+16x^2+26x+22}{x^3+3x^2+7x+5} dx &= \frac{1}{2}x^2+2x+2 \log|x+1| \\ &+ \frac{1}{2} \log\left[\left(\frac{x+1}{2}\right)^2+1\right] + \frac{1}{2} \arctan\left(\frac{x+1}{2}\right) + C \end{aligned}$$

The best thing after working through a few a nice long examples is to do another nice long example — it is excellent practice¹⁰. We recommend that the reader attempt the problem before reading through our solution.

✓ **Example 1.10.4** $\int \frac{4x^3+23x^2+45x+27}{x^3+5x^2+8x+4} dx$

In this example, we integrate $\frac{N(x)}{D(x)} = \frac{4x^3+23x^2+45x+27}{x^3+5x^2+8x+4}$.

- *Step 1.* The degree of the numerator $N(x)$ is equal to the degree of the denominator $D(x)$, so the first step to write $\frac{N(x)}{D(x)}$ in the form

$$\frac{N(x)}{D(x)} = P(x) + \frac{R(x)}{D(x)}$$

with $P(x)$ being a polynomial (which should be of degree 0, i.e. just a constant) and $R(x)$ being a polynomial of degree strictly smaller than the degree of $D(x)$. By long division



so

$$\frac{4x^3 + 23x^2 + 45x + 27}{x^3 + 5x^2 + 8x + 4} = 4 + \frac{3x^2 + 13x + 11}{x^3 + 5x^2 + 8x + 4}$$

- *Step 2.* The second step is to factorise $D(x) = x^3 + 5x^2 + 8x + 4$.
 - To start, we'll try and guess an integer root. Any integer root of $D(x)$ must divide the constant term, 4, exactly. Only $\pm 1, \pm 2, \pm 4$ can be integer roots of $x^3 + 5x^2 + 8x + 4$.
 - We test to see if ± 1 are roots.

$$\begin{aligned} D(1) &= (1)^3 + 5(1)^2 + 8(1) + 4 \neq 0 && \Rightarrow x = 1 \text{ is not a root} \\ D(-1) &= (-1)^3 + 5(-1)^2 + 8(-1) + 4 = 0 && \Rightarrow x = -1 \text{ is a root} \end{aligned}$$

So $(x + 1)$ must divide $x^3 + 5x^2 + 8x + 4$ exactly.

- By long division



so

$$\begin{aligned} x^3 + 5x^2 + 8x + 4 &= (x + 1)(x^2 + 4x + 4) \\ &= (x + 1)(x + 2)(x + 2) \end{aligned}$$

- Notice that we could have instead checked whether or not ± 2 are roots

$$\begin{aligned} D(2) &= (2)^3 + 5(2)^2 + 8(2) + 4 \neq 0 && \Rightarrow x = 2 \text{ is not a root} \\ D(-2) &= (-2)^3 + 5(-2)^2 + 8(-2) + 4 = 0 && \Rightarrow x = -2 \text{ is a root} \end{aligned}$$

We now know that both -1 and -2 are roots of $x^3 + 5x^2 + 8x + 4$ and hence both $(x + 1)$ and $(x + 2)$ are factors of $x^3 + 5x^2 + 8x + 4$. Because $x^3 + 5x^2 + 8x + 4$ is of degree three and the coefficient of x^3 is 1, we must have $x^3 + 5x^2 + 8x + 4 = (x + 1)(x + 2)(x + a)$ for some constant a . Multiplying out the right hand side shows that the constant term is $2a$. So $2a = 4$ and $a = 2$.

This is the end of step 2. We now know that

$$\frac{4x^3 + 23x^2 + 45x + 27}{x^3 + 5x^2 + 8x + 4} = 4 + \frac{3x^2 + 13x + 11}{(x + 1)(x + 2)^2}$$

- *Step 3.* The third step is to write $\frac{3x^2 + 13x + 11}{(x + 1)(x + 2)^2}$ in the form

$$\frac{3x^2 + 13x + 11}{(x + 1)(x + 2)^2} = \frac{A}{x + 1} + \frac{B}{x + 2} + \frac{C}{(x + 2)^2}$$

for some constants A, B and C .

Note that there are two terms on the right hand arising from the factor $(x + 2)^2$. One has denominator $(x + 2)$ and one has denominator $(x + 2)^2$. More generally, for each factor $(x + a)^n$ in the denominator of the rational function on the left hand side, we include

$$\frac{A_1}{x + a} + \frac{A_2}{(x + a)^2} + \dots + \frac{A_n}{(x + a)^n}$$

in the partial fraction decomposition on the right hand side¹¹.

To determine the values of the constants A, B, C , we put the right hand side back over the common denominator $(x + 1)(x + 2)^2$.

$$\begin{aligned} \frac{3x^2 + 13x + 11}{(x + 1)(x + 2)^2} &= \frac{A}{x + 1} + \frac{B}{x + 2} + \frac{C}{(x + 2)^2} \\ &= \frac{A(x + 2)^2 + B(x + 1)(x + 2) + C(x + 1)}{(x + 1)(x + 2)^2} \end{aligned}$$

The fraction on the far left is the same as the fraction on the far right if and only if their numerators are the same.

$$3x^2 + 13x + 11 = A(x+2)^2 + B(x+1)(x+2) + C(x+1)$$

As in the previous examples, there are a couple of different ways to determine the values of A , B and C from this equation.

- *Step 3 — Algebra Method.* The conceptually clearest procedure is to write the right hand side as a polynomial in standard form (i.e. collect up all x^2 terms, all x terms and all constant terms)

$$3x^2 + 13x + 11 = (A+B)x^2 + (4A+3B+C)x + (4A+2B+C)$$

For these two polynomials to be the same, the coefficient of x^2 on the left hand side and the coefficient of x^2 on the right hand side must be the same. Similarly the coefficients of x^1 and the coefficients of x^0 (i.e. the constant terms) must match. This gives us a system of three equations,

$$A + B = 3 \quad 4A + 3B + C = 13 \quad 4A + 2B + C = 11$$

in the three unknowns A, B, C . We can solve this system by

- using the first equation, namely $A + B = 3$, to determine A in terms of B : $A = 3 - B$.
- Substituting this into the remaining equations eliminates the A , leaving two equations in the two unknown B, C .

$$4(3 - B) + 3B + C = 13 \quad 4(3 - B) + 2B + C = 11$$

or

$$-B + C = 1 \quad -2B + C = -1$$

- We can now solve the first of these equations, namely $-B + C = 1$, for B in terms of C , giving $B = C - 1$.
- Substituting this into the last equation, namely $-2B + C = -1$, gives $-2(C - 1) + C = -1$ which is easily solved to give
- $C = 3$, and then $B = C - 1 = 2$ and then $A = 3 - B = 1$.

Hence

$$\begin{aligned} \frac{4x^3 + 23x^2 + 45x + 27}{x^3 + 5x^2 + 8x + 4} &= 4 + \frac{3x^2 + 13x + 11}{(x+1)(x+2)^2} \\ &= 4 + \frac{1}{x+1} + \frac{2}{x+2} + \frac{3}{(x+2)^2} \end{aligned}$$

- *Step 3 — Sneaky Method.* The second, sneakier, method for finding A, B and C exploits the fact that $3x^2 + 13x + 11 = A(x+2)^2 + B(x+1)(x+2) + C(x+1)$ must be true for all values of x . In particular, it must be true for $x = -1$. When $x = -1$, the factor $(x+1)$ multiplying B and C is exactly zero. So B and C disappear from the equation, leaving us with an easy equation to solve for A :

$$\begin{aligned} 3x^2 + 13x + 11 \Big|_{x=-1} &= \left[A(x+2)^2 + B(x+1)(x+2) + C(x+1) \right]_{x=-1} \\ &\implies 1 = A \end{aligned}$$

Sub this value of A back in and simplify.

$$\begin{aligned} 3x^2 + 13x + 11 &= (1)(x+2)^2 + B(x+1)(x+2) + C(x+1) \\ 2x^2 + 9x + 7 &= B(x+1)(x+2) + C(x+1) \\ &= (xB + 2B + C)(x+1) \end{aligned}$$

Since $(x+1)$ is a factor on the right hand side, it must also be a factor on the left hand side.

$$\begin{aligned} (2x+7)(x+1) &= (xB + 2B + C)(x+1) \\ \implies (2x+7) &= (xB + 2B + C) \end{aligned}$$

For the coefficients of x to match, B must be 2. For the constant terms to match, $2B + C$ must be 7, so C must be 3. Hence we again have

$$\begin{aligned} \frac{4x^3 + 23x^2 + 45x + 27}{x^3 + 5x^2 + 8x + 4} &= 4 + \frac{3x^2 + 13x + 11}{(x+1)(x+2)^2} \\ &= 4 + \frac{1}{x+1} + \frac{2}{x+2} + \frac{3}{(x+2)^2} \end{aligned}$$

- *Step 4.* The final step is to integrate

$$\begin{aligned} &\int \frac{4x^3 + 23x^2 + 45x + 27}{x^3 + 5x^2 + 8x + 4} dx \\ &= \int 4 dx + \int \frac{1}{x+1} dx + \int \frac{2}{x+2} dx + \int \frac{3}{(x+2)^2} dx \\ &= 4x + \log|x+1| + 2 \log|x+2| - \frac{3}{x+2} + C \end{aligned}$$

The method of partial fractions is not just confined to the problem of integrating rational functions. There are other integrals — such as $\int \sec x dx$ and $\int \sec^3 x dx$ — that can be transformed (via substitutions) into integrals of rational functions. We encountered both of these integrals in Sections 1.8 and 1.9 on trigonometric integrals and substitutions.

✓ Example 1.10.5 $\int \sec x dx$

Solution

In this example, we integrate $\sec x$. It is not yet clear what this integral has to do with partial fractions. To get to a partial fractions computation, we first make one of our old substitutions.

$$\begin{aligned} \int \sec x dx &= \int \frac{1}{\cos x} dx && \text{massage the expression a little} \\ &= \int \frac{\cos x}{\cos^2 x} dx && \text{substitute } u = \sin x, du = \cos x dx \\ &= - \int \frac{du}{u^2 - 1} && \text{and use } \cos^2 x = 1 - \sin^2 x = 1 - u^2 \end{aligned}$$

So we now have to integrate $\frac{1}{u^2-1}$, which is a rational function of u , and so is perfect for partial fractions.

- *Step 1.* The degree of the numerator, 1, is zero, which is strictly smaller than the degree of the denominator, $u^2 - 1$, which is two. So the first step is skipped.
- *Step 2.* The second step is to factor the denominator:

$$u^2 - 1 = (u - 1)(u + 1)$$

- *Step 3.* The third step is to write $\frac{1}{u^2-1}$ in the form

$$\frac{1}{u^2 - 1} = \frac{1}{(u - 1)(u + 1)} = \frac{A}{u - 1} + \frac{B}{u + 1}$$

for some constants A and B .

- *Step 3 — Sneaky Method.*
 - Multiply through by the denominator to get

$$1 = A(u + 1) + B(u - 1)$$

This equation must be true for all u .

- If we now set $u = 1$ then we eliminate B from the equation leaving us with

$$1 = 2A \quad \text{so } A = \frac{1}{2}.$$

- Similarly, if we set $u = -1$ then we eliminate A , leaving

$$1 = -2B \quad \text{which implies } B = -\frac{1}{2}.$$

We have now found that $A = \frac{1}{2}$, $B = -\frac{1}{2}$, so

$$\frac{1}{u^2 - 1} = \frac{1}{2} \left[\frac{1}{u - 1} - \frac{1}{u + 1} \right].$$

- It is always a good idea to check our work.

$$\frac{\frac{1}{2}}{u - 1} + \frac{-\frac{1}{2}}{u + 1} = \frac{\frac{1}{2}(u + 1) - \frac{1}{2}(u - 1)}{(u - 1)(u + 1)} = \frac{1}{(u - 1)(u + 1)} \checkmark$$

- *Step 4.* The final step is to integrate.

$$\begin{aligned} \int \sec x \, dx &= - \int \frac{du}{u^2 - 1} && \text{after substitution} \\ &= -\frac{1}{2} \int \frac{du}{u - 1} + \frac{1}{2} \int \frac{du}{u + 1} && \text{partial fractions} \\ &= -\frac{1}{2} \log |u - 1| + \frac{1}{2} \log |u + 1| + C \\ &= -\frac{1}{2} \log |\sin(x) - 1| + \frac{1}{2} \log |\sin(x) + 1| + C && \text{rearrange a little} \\ &= \frac{1}{2} \log \left| \frac{1 + \sin x}{1 - \sin x} \right| + C \end{aligned}$$

Notice that since $-1 \leq \sin x \leq 1$, we are free to drop the absolute values in the last line if we wish.

Another example in the same spirit, though a touch harder. Again, we saw this problem in Section 1.8 and 1.9.

✓ Example 1.10.6 $\int \sec^3 x \, dx$

Solution

- We'll start by converting it into the integral of a rational function using the substitution $u = \sin x$, $du = \cos x \, dx$.

$$\int \sec^3 x \, dx = \int \frac{1}{\cos^3 x} \, dx = \int \frac{\cos x}{\cos^4 x} \, dx = \int \frac{\cos x}{(1 - \sin^2 x)^2} \, dx$$
- We could now find the partial fraction decomposition of the integrand $\frac{1}{[1 - u^2]^2}$ by executing the usual four steps. But it is easier to use

$$\frac{1}{u^2 - 1} = \frac{1}{2} \left[\frac{1}{u - 1} - \frac{1}{u + 1} \right]$$

which we worked out in Example 1.10.5 above.

- Squaring this gives

$$\begin{aligned} \frac{1}{[1 - u^2]^2} &= \frac{1}{4} \left[\frac{1}{u - 1} - \frac{1}{u + 1} \right]^2 \\ &= \frac{1}{4} \left[\frac{1}{(u - 1)^2} - \frac{2}{(u - 1)(u + 1)} + \frac{1}{(u + 1)^2} \right] \\ &= \frac{1}{4} \left[\frac{1}{(u - 1)^2} - \frac{1}{u - 1} + \frac{1}{u + 1} + \frac{1}{(u + 1)^2} \right] \end{aligned}$$

where we have again used $\frac{1}{u^2 - 1} = \frac{1}{2} \left[\frac{1}{u - 1} - \frac{1}{u + 1} \right]$ in the last step.

- It only remains to do the integrals and simplify.

$$\begin{aligned}
 \int \sec^3 x \, dx &= \frac{1}{4} \int \left[\frac{1}{(u-1)^2} - \frac{1}{u-1} + \frac{1}{u+1} + \frac{1}{(u+1)^2} \right] du \\
 &= \frac{1}{4} \left[-\frac{1}{u-1} - \log |u-1| + \log |u+1| - \frac{1}{u+1} \right] + C \\
 &\quad \text{group carefully} \\
 &= \frac{-1}{4} \left[\frac{1}{u-1} + \frac{1}{u+1} \right] + \frac{1}{4} \left[\log |u+1| - \log |u-1| \right] + C \\
 &\quad \text{sum carefully} \\
 &= -\frac{1}{4} \frac{2u}{u^2-1} + \frac{1}{4} \log \left| \frac{u+1}{u-1} \right| + C \\
 &\quad \text{clean up} \\
 &= \frac{1}{2} \frac{u}{1-u^2} + \frac{1}{4} \log \left| \frac{u+1}{u-1} \right| + C \\
 &\quad \text{put } u = \sin x \\
 &= \frac{1}{2} \frac{\sin x}{\cos^2 x} + \frac{1}{4} \log \left| \frac{\sin x + 1}{\sin x - 1} \right| + C
 \end{aligned}$$

The form of partial fraction decompositions

In the examples above we used the partial fractions method to decompose rational functions into easily integrated pieces. Each of those examples was quite involved and we had to spend quite a bit of time factoring and doing long division. The key step in each of the computations was Step 3 — in that step we decomposed the rational function $\frac{N(x)}{D(x)}$ (or $\frac{R(x)}{D(x)}$), for which the degree of the numerator is strictly smaller than the degree of the denominator, into a sum of particularly simple rational functions, like $\frac{A}{x-a}$. We did not, however, give a systematic description of those decompositions.

In this subsection we fill that gap by describing the general¹² form of partial fraction decompositions. The justification of these forms is not part of the course, but the interested reader is invited to read the next (optional) subsection where such justification is given. In the following it is assumed that

- $N(x)$ and $D(x)$ are polynomials with the degree of $N(x)$ strictly smaller than the degree of $D(x)$.
- K is a constant.
- a_1, a_2, \dots, a_j are all different numbers.
- m_1, m_2, \dots, m_j , and n_1, n_2, \dots, n_k are all strictly positive integers.
- $x^2 + b_1x + c_1, x^2 + b_2x + c_2, \dots, x^2 + b_kx + c_k$ are all different.

Simple linear factor case

If the denominator $D(x) = K(x - a_1)(x - a_2) \cdots (x - a_j)$ is a product of j different linear factors, then

Equation 1.10.7

$$\frac{N(x)}{D(x)} = \frac{A_1}{x - a_1} + \frac{A_2}{x - a_2} + \cdots + \frac{A_j}{x - a_j}$$

We can then integrate each term

$$\int \frac{A}{x - a} \, dx = A \log |x - a| + C.$$

General linear factor case

If the denominator $D(x) = K(x - a_1)^{m_1}(x - a_2)^{m_2} \cdots (x - a_j)^{m_j}$ then

 Claim 1.10.8

$$\begin{aligned} \frac{N(x)}{D(x)} &= \frac{A_{1,1}}{x-a_1} + \frac{A_{1,2}}{(x-a_1)^2} + \cdots + \frac{A_{1,m_1}}{(x-a_1)^{m_1}} \\ &+ \frac{A_{2,1}}{x-a_2} + \frac{A_{2,2}}{(x-a_2)^2} + \cdots + \frac{A_{2,m_2}}{(x-a_2)^{m_2}} + \cdots \\ &+ \frac{A_{j,1}}{x-a_j} + \frac{A_{j,2}}{(x-a_j)^2} + \cdots + \frac{A_{j,m_j}}{(x-a_j)^{m_j}} \end{aligned}$$

Notice that we could rewrite each line as

$$\begin{aligned} \frac{A_1}{x-a} + \frac{A_2}{(x-a)^2} + \cdots + \frac{A_m}{(x-a)^m} &= \frac{A_1(x-a)^{m-1} + A_2(x-a)^{m-2} + \cdots + A_m}{(x-a)^m} \\ &= \frac{B_1x^{m-1} + B_2x^{m-2} + \cdots + B_m}{(x-a)^m} \end{aligned}$$

which is a polynomial whose degree, $m-1$, is strictly smaller than that of the denominator $(x-a)^m$. But the form of Equation 1.10.8 is preferable because it is easier to integrate.

$$\begin{aligned} \int \frac{A}{x-a} dx &= A \log|x-a| + C \\ \int \frac{A}{(x-a)^k} dx &= -\frac{1}{k-1} \cdot \frac{A}{(x-a)^{k-1}} \quad \text{provided } k > 1. \end{aligned}$$

Simple linear and quadratic factor case


If $D(x) = K(x-a_1) \cdots (x-a_j)(x^2+b_1x+c_1) \cdots (x^2+b_kx+c_k)$ then

 Claim 1.10.9

$$\frac{N(x)}{D(x)} = \frac{A_1}{x-a_1} + \cdots + \frac{A_j}{x-a_j} + \frac{B_1x+C_1}{x^2+b_1x+c_1} + \cdots + \frac{B_kx+C_k}{x^2+b_kx+c_k}$$

Note that the numerator of each term on the right hand side has degree one smaller than the degree of the denominator.

The quadratic terms $\frac{Bx+C}{x^2+bx+c}$ are integrated in a two-step process that is best illustrated with a simple example (see also Example 1.10.3 above).

 Example 1.10.10 $\int \frac{2x+7}{x^2+4x+13} dx$

Solution

- Start by completing the square in the denominator:

$$\begin{aligned} x^2 + 4x + 13 &= (x+2)^2 + 9 && \text{and thus} \\ \frac{2x+7}{x^2+4x+13} &= \frac{2x+7}{(x+2)^2 + 3^2} \end{aligned}$$

- Now set $y = (x+2)/3$, $dy = \frac{1}{3} dx$, or equivalently $x = 3y - 2$, $dx = 3 dy$:

$$\begin{aligned} \int \frac{2x+7}{x^2+4x+13} dx &= \int \frac{2x+7}{(x+2)^2+3^2} dx \\ &= \int \frac{6y-4+7}{3^2y^2+3^2} \cdot 3 dy \\ &= \int \frac{6y+3}{3(y^2+1)} dy \\ &= \int \frac{2y+1}{y^2+1} dy \end{aligned}$$

Notice that we chose 3 in $y = (x+2)/3$ precisely to transform the denominator into the form y^2+1 .

- Now almost always the numerator will be a linear polynomial of y and we decompose as follows

$$\begin{aligned} \int \frac{2x+7}{x^2+4x+13} dx &= \int \frac{2y+1}{y^2+1} dy \\ &= \int \frac{2y}{y^2+1} dy + \int \frac{1}{y^2+1} dy \\ &= \log|y^2+1| + \arctan y + C \\ &= \log\left|\left(\frac{x+2}{3}\right)^2+1\right| + \arctan\left(\frac{x+2}{3}\right) + C \end{aligned}$$

Optional — General linear and quadratic factor case

If $D(x) = K(x-a_1)^{m_1} \cdots (x-a_j)^{m_j} (x^2+b_1x+c_1)^{n_1} \cdots (x^2+b_kx+c_k)^{n_k}$

Claim 1.10.11

$$\begin{aligned} \frac{N(x)}{D(x)} &= \frac{A_{1,1}}{x-a_1} + \frac{A_{1,2}}{(x-a_1)^2} + \cdots + \frac{A_{1,m_1}}{(x-a_1)^{m_1}} + \cdots \\ &+ \frac{A_{j,1}}{x-a_j} + \frac{A_{j,2}}{(x-a_j)^2} + \cdots + \frac{A_{j,m_j}}{(x-a_j)^{m_j}} \\ &+ \frac{B_{1,1}x+C_{1,1}}{x^2+b_1x+c_1} + \frac{B_{1,2}x+C_{1,2}}{(x^2+b_1x+c_1)^2} + \cdots + \frac{B_{1,n_1}x+C_{1,n_1}}{(x^2+b_1x+c_1)^{n_1}} + \cdots \\ &+ \frac{B_{k,1}x+C_{k,1}}{x^2+b_kx+c_k} + \frac{B_{k,2}x+C_{k,2}}{(x^2+b_kx+c_k)^2} + \cdots + \frac{B_{k,n_k}x+C_{k,n_k}}{(x^2+b_kx+c_k)^{n_k}} \end{aligned}$$

We have already seen how to integrate the simple and general linear terms, and the simple quadratic terms. Integrating general quadratic terms is not so straightforward.

Example 1.10.12 $\int \frac{dx}{(x^2+1)^n}$

This example is not so easy, so it should definitely be considered optional.

Solution

In what follows write

$$I_n = \int \frac{dx}{(x^2+1)^n}.$$

- When $n = 1$ we know that

$$\int \frac{dx}{x^2+1} = \arctan x + C$$

- Now assume that $n > 1$, then

$$\begin{aligned} \int \frac{1}{(x^2+1)^n} dx &= \int \frac{(x^2+1-x^2)}{(x^2+1)^n} dx && \text{sneaky} \\ &= \int \frac{1}{(x^2+1)^{n-1}} dx - \int \frac{x^2}{(x^2+1)^n} dx \\ &= I_{n-1} - \int \frac{x^2}{(x^2+1)^n} dx \end{aligned}$$

So we can write I_n in terms of I_{n-1} and this second integral.

- We can use integration by parts to compute the second integral:

$$\int \frac{x^2}{(x^2+1)^n} dx = \int \frac{x}{2} \cdot \frac{2x}{(x^2+1)^n} dx \quad \text{sneaky}$$

We set $u = x/2$ and $dv = \frac{2x}{(x^2+1)^n} dx$, which gives $du = \frac{1}{2} dx$ and $v = -\frac{1}{n-1} \cdot \frac{1}{(x^2+1)^{n-1}}$. You can check v by differentiating. Integration by parts gives

$$\begin{aligned} \int \frac{x}{2} \cdot \frac{2x}{(x^2+1)^n} dx \\ &= -\frac{x}{2(n-1)(x^2+1)^{n-1}} + \int \frac{dx}{2(n-1)(x^2+1)^{n-1}} \\ &= -\frac{x}{2(n-1)(x^2+1)^{n-1}} + \frac{1}{2(n-1)} \cdot I_{n-1} \end{aligned}$$

- Now put everything together:

$$\begin{aligned} I_n &= \int \frac{1}{(x^2+1)^n} dx \\ &= I_{n-1} + \frac{x}{2(n-1)(x^2+1)^{n-1}} - \frac{1}{2(n-1)} \cdot I_{n-1} \\ &= \frac{2n-3}{2(n-1)} I_{n-1} + \frac{x}{2(n-1)(x^2+1)^{n-1}} \end{aligned}$$

- We can then use this recurrence to write down I_n for the first few n :

$$\begin{aligned} I_2 &= \frac{1}{2} I_1 + \frac{x}{2(x^2+1)} + C \\ &= \frac{1}{2} \arctan x + \frac{x}{2(x^2+1)} \\ I_3 &= \frac{3}{4} I_2 + \frac{x}{4(x^2+1)^2} \\ &= \frac{3}{8} \arctan x + \frac{3x}{8(x^2+1)} + \frac{x}{4(x^2+1)^2} + C \\ I_4 &= \frac{5}{6} I_3 + \frac{x}{6(x^2+1)^3} \\ &= \frac{5}{16} \arctan x + \frac{5x}{16(x^2+1)} + \frac{5x}{24(x^2+1)^2} + \frac{x}{6(x^2+1)^3} + C \end{aligned}$$

and so forth. You can see why partial fraction questions involving denominators with repeated quadratic factors do not often appear on exams.

Optional — Justification of the partial fraction decompositions

We will now see the justification for the form of the partial fraction decompositions. We will only consider the case in which the denominator has only linear factors. The arguments when there are quadratic factors too are similar ¹³.

Simple linear factor case

In the most common partial fraction decomposition, we split up

$$\frac{N(x)}{(x - a_1) \times \cdots \times (x - a_d)}$$

into a sum of the form

$$\frac{A_1}{x - a_1} + \cdots + \frac{A_d}{x - a_d}$$

We now show that this decomposition can always be achieved, under the assumptions that the a_i 's are all different and $N(x)$ is a polynomial of degree at most $d - 1$. To do so, we shall repeatedly apply the following Lemma.

Lemma 1.10.13

Let $N(x)$ and $D(x)$ be polynomials of degree n and d respectively, with $n \leq d$. Suppose that a is NOT a zero of $D(x)$. Then there is a polynomial $P(x)$ of degree $p < d$ and a number A such that

$$\frac{N(x)}{D(x)(x - a)} = \frac{P(x)}{D(x)} + \frac{A}{x - a}$$

Proof

- To save writing, let $z = x - a$. We then write $\tilde{N}(z) = N(z + a)$ and $\tilde{D}(z) = D(z + a)$, which are again polynomials of degree n and d respectively. We also know that $\tilde{D}(0) = D(a) \neq 0$.
- In order to complete the proof we need to find a polynomial $\tilde{P}(z)$ of degree $p < d$ and a number A such that

$$\frac{\tilde{N}(z)}{\tilde{D}(z)z} = \frac{\tilde{P}(z)}{\tilde{D}(z)} + \frac{A}{z} = \frac{\tilde{P}(z)z + A\tilde{D}(z)}{\tilde{D}(z)z}$$

or equivalently, such that

$$\tilde{P}(z)z + A\tilde{D}(z) = \tilde{N}(z).$$

- Now look at the polynomial on the left hand side. Every term in $\tilde{P}(z)z$, has at least one power of z . So the constant term on the left hand side is exactly the constant term in $A\tilde{D}(z)$, which is equal to $A\tilde{D}(0)$. The constant term on the right hand side is equal to $\tilde{N}(0)$. So the constant terms on the left and right hand sides are the same if we choose $A = \frac{\tilde{N}(0)}{\tilde{D}(0)}$. Recall that $\tilde{D}(0)$ cannot be zero, so A is well defined.
- Now move $A\tilde{D}(z)$ to the right hand side.

$$\tilde{P}(z)z = \tilde{N}(z) - A\tilde{D}(z)$$

The constant terms in $\tilde{N}(z)$ and $A\tilde{D}(z)$ are the same, so the right hand side contains no constant term and the right hand side is of the form $\tilde{N}_1(z)z$ for some polynomial $\tilde{N}_1(z)$.

- Since $\tilde{N}(z)$ is of degree at most d and $A\tilde{D}(z)$ is of degree exactly d , \tilde{N}_1 is a polynomial of degree $d - 1$. It now suffices to choose $\tilde{P}(z) = \tilde{N}_1(z)$.

Now back to

$$\frac{N(x)}{(x - a_1) \times \cdots \times (x - a_d)}$$

Apply Lemma 1.10.13, with $D(x) = (x - a_2) \times \cdots \times (x - a_d)$ and $a = a_1$. It says

$$\frac{N(x)}{(x - a_1) \times \cdots \times (x - a_d)} = \frac{A_1}{x - a_1} + \frac{P(x)}{(x - a_2) \times \cdots \times (x - a_d)}$$

for some polynomial P of degree at most $d - 2$ and some number A_1 .

Apply Lemma 1.10.13 a second time, with $D(x) = (x - a_3) \times \cdots \times (x - a_d)$, $N(x) = P(x)$ and $a = a_2$. It says

$$\frac{P(x)}{(x - a_2) \times \cdots \times (x - a_d)} = \frac{A_2}{x - a_2} + \frac{Q(x)}{(x - a_3) \times \cdots \times (x - a_d)}$$

for some polynomial Q of degree at most $d - 3$ and some number A_2 .

At this stage, we know that

$$\frac{N(x)}{(x - a_1) \times \cdots \times (x - a_d)} = \frac{A_1}{x - a_1} + \frac{A_2}{x - a_2} + \frac{Q(x)}{(x - a_3) \times \cdots \times (x - a_d)}$$

If we just keep going, repeatedly applying Lemma 1, we eventually end up with

$$\frac{N(x)}{(x - a_1) \times \cdots \times (x - a_d)} = \frac{A_1}{x - a_1} + \cdots + \frac{A_d}{x - a_d}$$

as required.

The general case with linear factors

Now consider splitting

$$\frac{N(x)}{(x - a_1)^{n_1} \times \cdots \times (x - a_d)^{n_d}}$$

into a sum of the form¹⁴

$$\left[\frac{A_{1,1}}{x - a_1} + \cdots + \frac{A_{1,n_1}}{(x - a_1)^{n_1}} \right] + \cdots + \left[\frac{A_{d,1}}{x - a_d} + \cdots + \frac{A_{d,n_d}}{(x - a_d)^{n_d}} \right]$$

We now show that this decomposition can always be achieved, under the assumptions that the a_i 's are all different and $N(x)$ is a polynomial of degree at most $n_1 + \cdots + n_d - 1$. To do so, we shall repeatedly apply the following Lemma.

Lemma 1.10.14

Let $N(x)$ and $D(x)$ be polynomials of degree n and d respectively, with $n < d + m$. Suppose that a is NOT a zero of $D(x)$. Then there is a polynomial $P(x)$ of degree $p < d$ and numbers A_1, \dots, A_m such that

$$\frac{N(x)}{D(x)(x - a)^m} = \frac{P(x)}{D(x)} + \frac{A_1}{x - a} + \frac{A_2}{(x - a)^2} + \cdots + \frac{A_m}{(x - a)^m}$$

Proof

- As we did in the proof of the previous lemma, we write $z = x - a$. Then $\tilde{N}(z) = N(z + a)$ and $\tilde{D}(z) = D(z + a)$ are polynomials of degree n and d respectively, $\tilde{D}(0) = D(a) \neq 0$.
- In order to complete the proof we have to find a polynomial $\tilde{P}(z)$ of degree $p < d$ and numbers A_1, \dots, A_m such that

$$\begin{aligned} \frac{\tilde{N}(z)}{\tilde{D}(z)z^m} &= \frac{\tilde{P}(z)}{\tilde{D}(z)} + \frac{A_1}{z} + \frac{A_2}{z^2} + \cdots + \frac{A_m}{z^m} \\ &= \frac{\tilde{P}(z)z^m + A_1z^{m-1}\tilde{D}(z) + A_2z^{m-2}\tilde{D}(z) + \cdots + A_m\tilde{D}(z)}{\tilde{D}(z)z^m} \end{aligned}$$

or equivalently, such that

$$\begin{aligned} \tilde{P}(z)z^m + A_1z^{m-1}\tilde{D}(z) + A_2z^{m-2}\tilde{D}(z) + \cdots + A_{m-1}z\tilde{D}(z) + A_m\tilde{D}(z) \\ = \tilde{N}(z) \end{aligned}$$

- Now look at the polynomial on the left hand side. Every single term on the left hand side, except for the very last one, $A_m \tilde{D}(z)$, has at least one power of z . So the constant term on the left hand side is exactly the constant term in $A_m \tilde{D}(z)$, which is equal to $A_m \tilde{D}(0)$. The constant term on the right hand side is equal to $\tilde{N}(0)$. So the constant terms on the left and right hand sides are the same if we choose $A_m = \frac{\tilde{N}(0)}{\tilde{D}(0)}$. Recall that $\tilde{D}(0) \neq 0$ so A_m is well defined.
- Now move $A_m \tilde{D}(z)$ to the right hand side.

$$\begin{aligned} \tilde{P}(z)z^m + A_1 z^{m-1} \tilde{D}(z) + A_2 z^{m-2} \tilde{D}(z) + \dots + A_{m-1} z \tilde{D}(z) \\ = \tilde{N}(z) - A_m \tilde{D}(z) \end{aligned}$$

The constant terms in $\tilde{N}(z)$ and $A_m \tilde{D}(z)$ are the same, so the right hand side contains no constant term and the right hand side is of the form $\tilde{N}_1(z)z$ with \tilde{N}_1 a polynomial of degree at most $d + m - 2$. (Recall that \tilde{N} is of degree at most $d + m - 1$ and \tilde{D} is of degree at most d .) Divide the whole equation by z to get

$$\tilde{P}(z)z^{m-1} + A_1 z^{m-2} \tilde{D}(z) + A_2 z^{m-3} \tilde{D}(z) + \dots + A_{m-1} \tilde{D}(z) = \tilde{N}_1(z).$$

- Now, we can repeat the previous argument. The constant term on the left hand side, which is exactly equal to $A_{m-1} \tilde{D}(0)$ matches the constant term on the right hand side, which is equal to $\tilde{N}_1(0)$ if we choose $A_{m-1} = \frac{\tilde{N}_1(0)}{\tilde{D}(0)}$. With this choice of A_{m-1}

$$\begin{aligned} \tilde{P}(z)z^{m-1} + A_1 z^{m-2} \tilde{D}(z) + A_2 z^{m-3} \tilde{D}(z) + \dots + A_{m-2} z \tilde{D}(z) \\ = \tilde{N}_1(z) - A_{m-1} \tilde{D}(z) = \tilde{N}_2(z)z \end{aligned}$$

with \tilde{N}_2 a polynomial of degree at most $d + m - 3$. Divide by z and continue.

- After m steps like this, we end up with

$$\tilde{P}(z)z = \tilde{N}_{m-1}(z) - A_1 \tilde{D}(z)$$

after having chosen $A_1 = \frac{\tilde{N}_{m-1}(0)}{\tilde{D}(0)}$.

- There is no constant term on the right side so that $\tilde{N}_{m-1}(z) - A_1 \tilde{D}(z)$ is of the form $\tilde{N}_m(z)z$ with \tilde{N}_m a polynomial of degree $d - 1$. Choosing $\tilde{P}(z) = \tilde{N}_m(z)$ completes the proof.

Now back to

$$\frac{N(x)}{(x - a_1)^{n_1} \times \dots \times (x - a_d)^{n_d}}$$

Apply Lemma 1.10.14, with $D(x) = (x - a_2)^{n_2} \times \dots \times (x - a_d)^{n_d}$, $m = n_1$ and $a = a_1$. It says

$$\begin{aligned} \frac{N(x)}{(x - a_1)^{n_1} \times \dots \times (x - a_d)^{n_d}} \\ = \frac{A_{1,1}}{x - a_1} + \frac{A_{1,2}}{(x - a_1)^2} + \dots + \frac{A_{1,n_1}}{(x - a_1)^{n_1}} + \frac{P(x)}{(x - a_2)^{n_2} \times \dots \times (x - a_d)^{n_d}} \end{aligned}$$

Apply Lemma 1.10.14 a second time, with $D(x) = (x - a_3)^{n_3} \times \dots \times (x - a_d)^{n_d}$, $N(x) = P(x)$, $m = n_2$ and $a = a_2$. And so on. Eventually, we end up with

$$\left[\frac{A_{1,1}}{x - a_1} + \dots + \frac{A_{1,n_1}}{(x - a_1)^{n_1}} \right] + \dots + \left[\frac{A_{d,1}}{x - a_d} + \dots + \frac{A_{d,n_d}}{(x - a_d)^{n_d}} \right]$$

which is exactly what we were trying to show.

Really Optional — The Fully General Case

We are now going to see that, in general, if $N(x)$ and $D(x)$ are polynomials with the degree of N being strictly smaller than the degree of D (which we'll denote $\deg(N) < \deg(D)$) and if

$$D(x) = K(x - a_1)^{m_1} \dots (x - a_j)^{m_j} (x^2 + b_1x + c_1)^{n_1} \dots (x^2 + b_kx + c_k)^{n_k} \quad (\star)$$

(with $b_\ell^2 - 4c_\ell < 0$ for all $1 \leq \ell \leq k$ so that no quadratic factor can be written as a product of linear factors with real coefficients) then there are real numbers $A_{i,j}, B_{i,j}, C_{i,j}$ such that

$$\begin{aligned} \frac{N(x)}{D(x)} &= \frac{A_{1,1}}{x-a_1} + \frac{A_{1,2}}{(x-a_1)^2} + \cdots + \frac{A_{1,m_1}}{(x-a_1)^{m_1}} + \cdots \\ &+ \frac{A_{j,1}}{x-a_j} + \frac{A_{j,2}}{(x-a_j)^2} + \cdots + \frac{A_{j,m_j}}{(x-a_j)^{m_j}} \\ &+ \frac{B_{1,1}x+C_{1,1}}{x^2+b_1x+c_1} + \frac{B_{1,2}x+C_{1,2}}{(x^2+b_1x+c_1)^2} + \cdots + \frac{B_{1,n_1}x+C_{1,n_1}}{(x^2+b_1x+c_1)^{n_1}} + \cdots \\ &+ \frac{B_{k,1}x+C_{k,1}}{x^2+b_kx+c_k} + \frac{B_{k,2}x+C_{k,2}}{(x^2+b_kx+c_k)^2} + \cdots + \frac{B_{k,n_k}x+C_{k,n_k}}{(x^2+b_kx+c_k)^{n_k}} \end{aligned}$$

This was Equation 1.10.11. We start with two simpler results, that we'll use repeatedly to get Equation 1.10.11. In the first simpler result, we consider the fraction $\frac{P(x)}{Q_1(x)Q_2(x)}$ with $P(x), Q_1(x)$ and $Q_2(x)$ being polynomials with real coefficients and we are going to assume that when $P(x), Q_1(x)$ and $Q_2(x)$ are factored as in (\star) , no two of them have a common linear or quadratic factor. As an example, no two of

$$\begin{aligned} P(x) &= 2(x-3)(x-4)(x^2+3x+3) \\ Q_1(x) &= 2(x-1)(x^2+2x+2) \\ Q_2(x) &= 2(x-2)(x^2+2x+3) \end{aligned}$$

have such a common factor. But, for

$$\begin{aligned} P(x) &= 2(x-3)(x-4)(x^2+x+1) \\ Q_1(x) &= 2(x-1)(x^2+2x+2) \\ Q_2(x) &= 2(x-2)(x^2+x+1) \end{aligned}$$

$P(x)$ and $Q_2(x)$ have the common factor x^2+x+1 .

Lemma 1.10.15

Let $P(x), Q_1(x)$ and $Q_2(x)$ be polynomials with real coefficients and with $\deg(P) < \deg(Q_1Q_2)$. Assume that no two of $P(x), Q_1(x)$ and $Q_2(x)$ have a common linear or quadratic factor. Then there are polynomials P_1, P_2 with $\deg(P_1) < \deg(Q_1), \deg(P_2) < \deg(Q_2)$, and

$$\frac{P(x)}{Q_1(x)Q_2(x)} = \frac{P_1(x)}{Q_1(x)} + \frac{P_2(x)}{Q_2(x)}$$

Proof

We are to find polynomials P_1 and P_2 that obey

$$P(x) = P_1(x)Q_2(x) + P_2(x)Q_1(x)$$

Actually, we are going to find polynomials p_1 and p_2 that obey

$$p_1(x)Q_1(x) + p_2(x)Q_2(x) = C \tag{**}$$

for some nonzero constant C , and then just multiply $(**)$ by $\frac{P(x)}{C}$. To find p_1, p_2 and C we are going to use something called the Euclidean algorithm. It is an algorithm¹⁵ that is used to efficiently find the greatest common divisors of two numbers. Because $Q_1(x)$ and $Q_2(x)$ have no common factors of degree 1 or 2, their "greatest common divisor" has degree 0, i.e. is a constant.

- The first step is to apply long division to $\frac{Q_1(x)}{Q_2(x)}$ to find polynomials $n_0(x)$ and $r_0(x)$ such that

$$\frac{Q_1(x)}{Q_2(x)} = n_0(x) + \frac{r_0(x)}{Q_2(x)} \quad \text{with } \deg(r_0) < \deg(Q_2)$$

or, equivalently,

$$Q_1(x) = n_0(x) Q_2(x) + r_0(x) \quad \text{with } \deg(r_0) < \deg(Q_2)$$

- The second step is to apply long division to $\frac{Q_2(x)}{r_0(x)}$ to find polynomials $n_1(x)$ and $r_1(x)$ such that

$$Q_2(x) = n_1(x) r_0(x) + r_1(x) \quad \text{with } \deg(r_1) < \deg(r_0) \text{ or } r_1(x) = 0$$

- The third step (assuming that $r_1(x)$ was not zero) is to apply long division to $\frac{r_0(x)}{r_1(x)}$ to find polynomials $n_2(x)$ and $r_2(x)$ such that

$$r_0(x) = n_2(x) r_1(x) + r_2(x) \quad \text{with } \deg(r_2) < \deg(r_1) \text{ or } r_2(x) = 0$$

- And so on.

As the degree of the remainder $r_i(x)$ decreases by at least one each time i is increased by one, the above iteration has to terminate with some $r_{\ell+1}(x) = 0$. That is, we choose ℓ to be index of the last nonzero remainder. Here is a summary of all of the long division steps.

$$\begin{aligned} Q_1(x) &= n_0(x) Q_2(x) + r_0(x) && \text{with } \deg(r_0) < \deg(Q_2) \\ Q_2(x) &= n_1(x) r_0(x) + r_1(x) && \text{with } \deg(r_1) < \deg(r_0) \\ r_0(x) &= n_2(x) r_1(x) + r_2(x) && \text{with } \deg(r_2) < \deg(r_1) \\ r_1(x) &= n_3(x) r_2(x) + r_3(x) && \text{with } \deg(r_3) < \deg(r_2) \\ &\vdots && \\ r_{\ell-2}(x) &= n_{\ell}(x) r_{\ell-1}(x) + r_{\ell}(x) && \text{with } \deg(r_{\ell}) < \deg(r_{\ell-1}) \\ r_{\ell-1}(x) &= n_{\ell+1}(x) r_{\ell}(x) + r_{\ell+1}(x) && \text{with } r_{\ell+1} = 0 \end{aligned}$$

Now we are going to take a closer look at all of the different remainders that we have generated.

- From first long division step, namely $Q_1(x) = n_0(x) Q_2(x) + r_0(x)$ we have that the remainder

$$r_0(x) = Q_1(x) - n_0(x) Q_2(x)$$

- From the second long division step, namely $Q_2(x) = n_1(x) r_0(x) + r_1(x)$ we have that the remainder

$$\begin{aligned} r_1(x) &= Q_2(x) - n_1(x) r_0(x) = Q_2(x) - n_1(x) [Q_1(x) - n_0(x) Q_2(x)] \\ &= A_1(x) Q_1(x) + B_1(x) Q_2(x) \end{aligned}$$

with $A_1(x) = -n_1(x)$ and $B_1(x) = 1 + n_0(x) n_1(x)$.

- From the third long division step (assuming that $r_1(x)$ was not zero), namely $r_0(x) = n_2(x) r_1(x) + r_2(x)$, we have that the remainder

$$\begin{aligned} r_2(x) &= r_0(x) - n_2(x) r_1(x) \\ &= [Q_1(x) - n_0(x) Q_2(x)] - n_2(x) [A_1(x) Q_1(x) + B_1(x) Q_2(x)] \\ &= A_2(x) Q_1(x) + B_2(x) Q_2(x) \end{aligned}$$

with $A_2(x) = 1 - n_2(x) A_1(x)$ and $B_2(x) = -n_0(x) - n_2(x) B_1(x)$.

- And so on. Continuing in this way, we conclude that the final nonzero remainder $r_{\ell}(x) = A_{\ell}(x) Q_1(x) + B_{\ell}(x) Q_2(x)$ for some polynomials A_{ℓ} and B_{ℓ} .

Now the last nonzero remainder $r_{\ell}(x)$ has to be a nonzero constant C because

- it is nonzero by the definition of $r_{\ell}(x)$ and
- if $r_{\ell}(x)$ were a polynomial of degree at least one, then
- so that $r_{\ell}(x)$ would be a common factor for $Q_1(x)$ and $Q_2(x)$, in contradiction to the hypothesis that no two of $P(x)$, $Q_1(x)$ and $Q_2(x)$ have a common linear or quadratic factor.

We now have that $A_{\ell}(x) Q_1(x) + B_{\ell}(x) Q_2(x) = r_{\ell}(x) = C$. Multiplying by $\frac{P(x)}{C}$ gives

$$\tilde{P}_2(x) Q_1(x) + \tilde{P}_1(x) Q_2(x) = P(x) \quad \text{or} \quad \frac{\tilde{P}_1(x)}{Q_1(x)} + \frac{\tilde{P}_2(x)}{Q_2(x)} = \frac{P(x)}{Q_1(x) Q_2(x)}$$

with $\tilde{P}_2(x) = \frac{P(x) A_\ell(x)}{C}$ and $\tilde{P}_1(x) = \frac{P(x) B_\ell(x)}{C}$. We're not quite done, because there is still the danger that $\deg(\tilde{P}_1) \geq \deg(Q_1)$ or $\deg(\tilde{P}_2) \geq \deg(Q_2)$. To deal with that possibility, we long divide $\frac{\tilde{P}_1(x)}{Q_1(x)}$ and call the remainder $P_1(x)$.

$$\frac{\tilde{P}_1(x)}{Q_1(x)} = N(x) + \frac{P_1(x)}{Q_1(x)} \quad \text{with } \deg(P_1) < \deg(Q_1)$$

Therefore we have that

$$\begin{aligned} \frac{P(x)}{Q_1(x) Q_2(x)} &= \frac{P_1(x)}{Q_1(x)} + N(x) + \frac{\tilde{P}_2(x)}{Q_2(x)} \\ &= \frac{P_1(x)}{Q_1(x)} + \frac{\tilde{P}_2(x) + N(x) Q_2(x)}{Q_2(x)} \end{aligned}$$

Denoting $P_2(x) = \tilde{P}_2(x) + N(x) Q_2(x)$ gives $\frac{P}{Q_1 Q_2} = \frac{P_1}{Q_1} + \frac{P_2}{Q_2}$ and since $\deg(P_1) < \deg(Q_1)$, the only thing left to prove is that $\deg(P_2) < \deg(Q_2)$. We assume that $\deg(P_2) \geq \deg(Q_2)$ and look for a contradiction. We have

$$\begin{aligned} \deg(P_2 Q_1) &\geq \deg(Q_1 Q_2) > \deg(P_1 Q_2) \\ \implies \deg(P) = \deg(P_1 Q_2 + P_2 Q_1) &= \deg(P_2 Q_1) \geq \deg(Q_1 Q_2) \end{aligned}$$

which contradicts the hypothesis that $\deg(P) < \deg(Q_1 Q_2)$ and the proof is complete.

For the second of the two simpler results, that we'll shortly use repeatedly to get Equation 1.10.11, we consider $\frac{P(x)}{(x-a)^m}$ and $\frac{P(x)}{(x^2+bx+c)^m}$.

Lemma 1.10.16

Let $m \geq 2$ be an integer, and let $Q(x)$ be either $x - a$ or $x^2 + bx + c$, with a , b and c being real numbers. Let $P(x)$ be a polynomial with real coefficients, which does not contain $Q(x)$ as a factor, and with $\deg(P) < \deg(Q^m) = m \deg(Q)$. Then, for each $1 \leq i \leq m$, there is a polynomial P_i with $\deg(P_i) < \deg(Q)$ or $P_i = 0$, such that

$$\frac{P(x)}{Q(x)^m} = \frac{P_1(x)}{Q(x)} + \frac{P_2(x)}{Q(x)^2} + \frac{P_3(x)}{Q(x)^3} + \cdots + \frac{P_{m-1}(x)}{Q(x)^{m-1}} + \frac{P_m(x)}{Q(x)^m}.$$

In particular, if $Q(x) = x - a$, then each $P_i(x)$ is just a constant A_i , and if $Q(x) = x^2 + bx + c$, then each $P_i(x)$ is a polynomial $B_i x + C_i$ of degree at most one.

Proof

We simply repeatedly use long division to get

$$\begin{aligned}
 \frac{P(x)}{Q(x)^m} &= \frac{P(x)}{Q(x)} \frac{1}{Q(x)^{m-1}} = \left\{ n_1(x) + \frac{r_1(x)}{Q(x)} \right\} \frac{1}{Q(x)^{m-1}} \\
 &= \frac{r_1(x)}{Q(x)^m} + \frac{n_1(x)}{Q(x)} \frac{1}{Q(x)^{m-2}} \\
 &= \frac{r_1(x)}{Q(x)^m} + \left\{ n_2(x) + \frac{r_2(x)}{Q(x)} \right\} \frac{1}{Q(x)^{m-2}} \\
 &= \frac{r_1(x)}{Q(x)^m} + \frac{r_2(x)}{Q(x)^{m-1}} + \frac{n_2(x)}{Q(x)} \frac{1}{Q(x)^{m-3}} \\
 &\vdots \\
 &= \frac{r_1(x)}{Q(x)^m} + \frac{r_2(x)}{Q(x)^{m-1}} + \cdots + \frac{r_{m-2}(x)}{Q(x)^3} + \frac{n_{m-2}(x)}{Q(x)} \frac{1}{Q(x)} \\
 &= \frac{r_1(x)}{Q(x)^m} + \frac{r_2(x)}{Q(x)^{m-1}} + \cdots + \frac{r_{m-2}(x)}{Q(x)^3} + \\
 &\quad \left\{ n_{m-1}(x) + \frac{r_{m-1}(x)}{Q(x)} \right\} \frac{1}{Q(x)} \\
 &= \frac{r_1(x)}{Q(x)^m} + \frac{r_2(x)}{Q(x)^{m-1}} + \cdots + \frac{r_{m-2}(x)}{Q(x)^3} + \frac{r_{m-1}(x)}{Q(x)^2} + \frac{n_{m-1}(x)}{Q(x)}
 \end{aligned}$$

By the rules of long division every $\deg(r_i) < \deg(Q)$. It is also true that the final numerator, n_{m-1} , has $\deg(n_{m-1}) < \deg(Q)$ — that is, we kept dividing by Q until the degree of the quotient was less than the degree of Q . To see this, note that $\deg(P) < m \deg(Q)$ and

$$\begin{aligned}
 \deg(n_1) &= \deg(P) - \deg(Q) \\
 \deg(n_2) &= \deg(n_1) - \deg(Q) = \deg(P) - 2 \deg(Q) \\
 &\vdots \\
 \deg(n_{m-1}) &= \deg(n_{m-2}) - \deg(Q) = \deg(P) - (m-1) \deg(Q) \\
 &< m \deg(Q) - (m-1) \deg(Q) \\
 &= \deg(Q)
 \end{aligned}$$

So, if $\deg(Q) = 1$, then $r_1, r_2, \dots, r_{m-1}, n_{m-1}$ are all real numbers, and if $\deg(Q) = 2$, then $r_1, r_2, \dots, r_{m-1}, n_{m-1}$ all have degree at most one.

We are now in a position to get Equation 1.10.11. We use $(*)$ to factor $^{16} D(x) = (x - a_1)^{m_1} Q_2(x)$ and use Lemma 1.10.15 to get

$$\frac{N(x)}{D(x)} = \frac{N(x)}{(x - a_1)^{m_1} Q_2(x)} = \frac{P_1(x)}{(x - a_1)^{m_1}} + \frac{P_2(x)}{Q_2(x)}$$

where $\deg(P_1) < m_1$, and $\deg(P_2) < \deg(Q_2)$. Then we use Lemma Lemma 1.10.16 to get

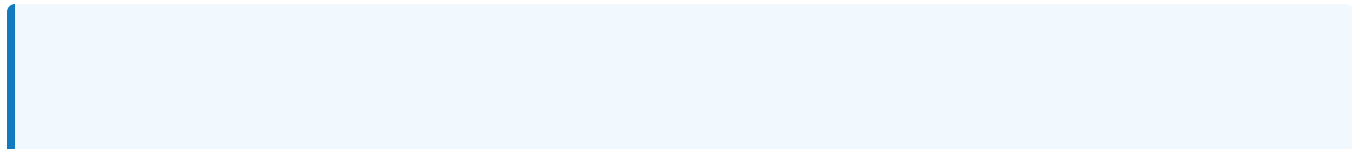
$$\frac{N(x)}{D(x)} = \frac{P_1(x)}{(x - a_1)^{m_1}} + \frac{P_2(x)}{Q_2(x)} = \frac{A_{1,1}}{x - a_1} + \frac{A_{1,2}}{(x - a_1)^2} + \cdots + \frac{A_{1,m_1}}{(x - a_1)^{m_1}} + \frac{P_2(x)}{Q_2(x)}$$

We continue working on $\frac{P_2(x)}{Q_2(x)}$ in this way, pulling off of the denominator one $(x - a_i)^{m_i}$ or one $(x^2 + b_i x + c_i)^{n_i}$ at a time, until we exhaust all of the factors in the denominator $D(x)$.

Exercises

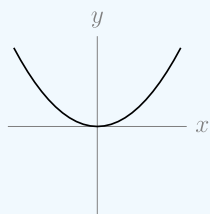
Recall that we are using $\log x$ to denote the logarithm of x with base e . In other courses it is often denoted $\ln x$.

Stage 1

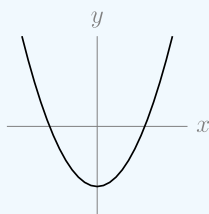


? 1

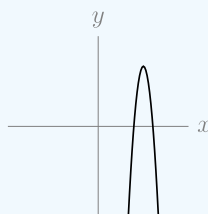
Below are the graphs of four different quadratic functions. For each quadratic function, decide whether it is: (i) irreducible, (ii) the product of two distinct linear factors, or (iii) the product of a repeated linear factor (and possibly a constant).



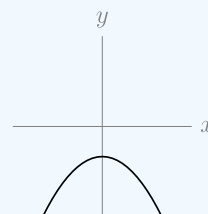
(a)



(b)



(c)



(d)

? 2 (*)

Write out the **general form** of the partial-fractions decomposition of $\frac{x^3 + 3}{(x^2 - 1)^2(x^2 + 1)}$. You need not determine the values of any of the coefficients.

? 3 (*)

Find the coefficient of $\frac{1}{x-1}$ in the partial fraction decomposition of $\frac{3x^3 - 2x^2 + 11}{x^2(x-1)(x+3)}$.

? 4

Re-write the following rational functions as the sum of a polynomial and a rational function whose numerator has a strictly smaller degree than its denominator. (Remember our method of partial fraction decomposition of a rational function only works when the degree of the numerator is strictly smaller than the degree of the denominator.)

1. $\frac{x^3 + 2x + 2}{x^2 + 1}$

2. $\frac{15x^4 + 6x^3 + 34x^2 + 4x + 20}{5x^2 + 2x + 8}$

3. $\frac{2x^5 + 9x^3 + 12x^2 + 10x + 30}{2x^2 + 5}$

? 5

Factor the following polynomials into linear and irreducible factors.

1. $5x^3 - 3x^2 - 10x + 6$

2. $x^4 - 3x^2 - 5$

3. $x^4 - 4x^3 - 10x^2 - 11x - 6$

4. $2x^4 + 12x^3 - x^2 - 52x + 15$

? 6

Here is a fact:

Suppose we have a rational function with a repeated linear factor $(ax + b)^n$ in the denominator, and the degree of the numerator is strictly less than the degree of the denominator. In the partial fraction decomposition, we can replace the terms

$$\frac{A_1}{ax+b} + \frac{A_2}{(ax+b)^2} + \frac{A_3}{(ax+b)^3} + \cdots + \frac{A_n}{(ax+b)^n}$$

with the single term

$$\frac{B_1 + B_2x + B_3x^2 + \cdots + B_nx^{n-1}}{(ax+b)^n}$$

and still be guaranteed to find a solution.

Why do we use the sum in (1), rather than the single term in (2), in partial fraction decomposition?

Stage 2

? 7 (*)

Evaluate $\int_1^2 \frac{dx}{x+x^2}$.

? 8 (*)

Calculate $\int \frac{1}{x^4+x^2} dx$.

? 9 (*)

Calculate $\int \frac{12x+4}{(x-3)(x^2+1)} dx$.

? 10 (*)

Evaluate the following indefinite integral using partial fraction:

$$F(x) = \int \frac{3x^2 - 4}{(x-2)(x^2+4)} dx.$$

? 11 (*)

Evaluate $\int \frac{x-13}{x^2-x-6} dx$.

? 12 (*)

Evaluate $\int \frac{5x+1}{x^2+5x+6} dx$.

? 13

Evaluate $\int \frac{5x^2-3x-1}{x^2-1} dx$.

? 14

Evaluate $\int \frac{4x^4 + 14x^2 + 2}{4x^4 + x^2} dx$.

? 15

Evaluate $\int \frac{x^2 + 2x - 1}{x^4 - 2x^3 + x^2} dx$.

? 16

Evaluate $\int \frac{3x^2 - 4x - 10}{2x^3 - x^2 - 8x + 4} dx$.

? 17

Evaluate $\int_0^1 \frac{10x^2 + 24x + 8}{2x^3 + 11x^2 + 6x + 5} dx$.

Stage 3

In Questions 18 and 19, we use partial fraction to find the antiderivatives of two important functions: cosecant, and cosecant cubed.

The purpose of performing a partial fraction decomposition is to manipulate an integrand into a form that is easily integrable. These “easily integrable” forms are rational functions whose denominator is a power of a linear function, or of an irreducible quadratic function. In Questions 20 through 23, we explore the integration of rational functions whose denominators involve irreducible quadratics.

In Questions 24 through 26, we use substitution to turn a non-rational integrand into a rational integrand, then evaluate the resulting integral using partial fraction. Till now, the partial fraction problems you've seen have all looked largely the same, but keep in mind that a partial fraction decomposition can be a small step in a larger problem.

? 18

Using the method of Example 1.10.5, integrate $\int \csc x dx$.

? 19

Using the method of Example 1.10.6, integrate $\int \csc^3 x dx$.

? 20

Evaluate $\int_1^2 \frac{3x^3 + 15x^2 + 35x + 10}{x^4 + 5x^3 + 10x^2} dx$.

? 21

Evaluate $\int \left(\frac{3}{x^2 + 2} + \frac{x - 3}{(x^2 + 2)^2} \right) dx$.

? 22

Evaluate $\int \frac{1}{(1+x^2)^3} dx$.

? 23

Evaluate $\int \left(3x + \frac{3x+1}{x^2+5} + \frac{3x}{(x^2+5)^2} \right) dx$.

? 24

Evaluate $\int \frac{\cos \theta}{3 \sin \theta + \cos^2 \theta - 3} d\theta$.

? 25

Evaluate $\int \frac{1}{e^{2t} + e^t + 1} dt$.

? 26

Evaluate $\int \sqrt{1+e^x} dx$ using partial fraction.

? 27 (*)

The region R is the portion of the first quadrant where $3 \leq x \leq 4$ and $0 \leq y \leq \frac{10}{\sqrt{25-x^2}}$.

1. Sketch the region R .
2. Determine the volume of the solid obtained by revolving R around the x -axis.
3. Determine the volume of the solid obtained by revolving R around the y -axis.

? 28

Find the area of the finite region bounded by the curves $y = \frac{4}{3+x^2}$, $y = \frac{2}{x(x+1)}$, $x = \frac{1}{4}$, and $x = 3$.

? 29

Let $F(x) = \int_1^x \frac{1}{t^2-9} dt$.

1. Give a formula for $F(x)$ that does not involve an integral.
2. Find $F'(x)$.

1. Recall that a rational function is the ratio of two polynomials.
2. The degree of a polynomial is the largest power of x . For example, the degree of $2x^3 + 4x^2 + 6x + 8$ is three.
3. We will soon get to an example (Example 1.10.2 in fact) in which the numerator degree is at least as large as the denominator degree — in that situation we have to extract a polynomial $P(x)$ before we can move on to step 2.
4. That is, we take the decomposed form and sum it back together.
5. Though, in fairness, we did step 3 twice — and that is the most tedious bit... Actually — sometimes factoring the denominator can be quite challenging. We'll consider this issue in more detail shortly.

6. One does not typically think of mathematics assignments or exams as nice kind places... The polynomials that appear in the “real world” are not so forgiving. Nature, red in tooth and claw — to quote Tennyson inappropriately (especially when this author doesn't know any other words from the poem).
7. Appendix A.16 contains several simple tricks for factoring polynomials. We recommend that you have a look at them.
8. To be precise, the quadratic equation $ax^2 + bx + c = 0$ has solutions $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. The term $b^2 - 4ac$ is called the discriminant and it tells us about the number of solutions. If the discriminant is positive then there are two real solutions. When it is zero, there is a single solution. And if it is negative, there is no real solutions (you need complex numbers to say more than this).
9. This same idea arose in Section 1.9. Given a quadratic written as $Q(x) = ax^2 + bx + c$ rewrite it as $Q(x) = a(x + d)^2 + e$. We can determine d and e by expanding and comparing coefficients of x :
 $ax^2 + bx + c = a(x^2 + 2dx + d^2) + e = ax^2 + 2dax + (e + ad^2)$. Hence $d = b/2a$ and $e = c - ad^2$.
10. At the risk of quoting Nietzsche, “That which does not kill us makes us stronger.” Though this author always preferred the logically equivalent contrapositive — “That which does not make us stronger will kill us.” However no one is likely to be injured by practicing partial fractions or looking up quotes on Wikipedia. Its also a good excuse to remind yourself of what a contrapositive is — though we will likely look at them again when we get to sequences and series.
11. This is justified in the (optional) subsection “Justification of the Partial Fraction Decompositions” below.
12. Well — not the completely general form, in the sense that we are not allowing the use of complex numbers. As a result we have to use both linear and quadratic factors in the denominator. If we could use complex numbers we would be able to restrict ourselves to linear factors.
13. In fact, quadratic factors are completely avoidable because, if we use complex numbers, then every polynomial can be written as a product of linear factors. This is the fundamental theorem of algebra.
14. If we allow ourselves to use complex numbers as roots, this is the general case. We don't need to consider quadratic (or higher) factors since all polynomials can be written as products of linear factors with complex coefficients.
15. It appears in Euclid's Elements, which was written about 300 BC, and it was probably known even before that.
16. This is assuming that there is at least one linear factor. If not, we factor $D(x) = (x^2 + b_1x + c_1)^{n_1} Q_2(x)$ instead.

This page titled [1.10: Partial Fractions](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.11: Numerical Integration

By now the reader will have come to appreciate that integration is generally quite a bit more difficult than differentiation. There are a great many simple-looking integrals, such as $\int e^{-x^2} dx$, that are either very difficult or even impossible to express in terms of standard functions¹. Such integrals are not merely mathematical curiosities, but arise very naturally in many contexts. For example, the error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

is extremely important in many areas of mathematics, and also in many practical applications of statistics.

In such applications we need to be able to evaluate this integral (and many others) at a given numerical value of x . In this section we turn to the problem of how to find (approximate) numerical values for integrals, without having to evaluate them algebraically. To develop these methods we return to Riemann sums and our geometric interpretation of the definite integral as the signed area.

We start by describing (and applying) three simple algorithms for generating, numerically, approximate values for the definite integral $\int_a^b f(x) dx$. In each algorithm, we begin in much the same way as we approached Riemann sums.

- We first select an integer $n > 0$, called the “number of steps”.
- We then divide the interval of integration, $a \leq x \leq b$, into n equal subintervals, each of length $\Delta x = \frac{b-a}{n}$. The first subinterval runs from $x_0 = a$ to $x_1 = a + \Delta x$. The second runs from x_1 to $x_2 = a + 2\Delta x$, and so on. The last runs from $x_{n-1} = b - \Delta x$ to $x_n = b$.



This splits the original integral into n pieces:

$$\int_a^b f(x) dx = \int_{x_0}^{x_1} f(x) dx + \int_{x_1}^{x_2} f(x) dx + \cdots + \int_{x_{n-1}}^{x_n} f(x) dx$$

Each subintegral $\int_{x_{j-1}}^{x_j} f(x) dx$ is approximated by the area of a simple geometric figure. The three algorithms we consider approximate the area by rectangles, trapezoids and parabolas (respectively).



We will explain these rules in detail below, but we give a brief overview here:

1. The midpoint rule approximates each subintegral by the area of a rectangle of height given by the value of the function at the midpoint of the subinterval

$$\int_{x_{j-1}}^{x_j} f(x) dx \approx f\left(\frac{x_{j-1} + x_j}{2}\right) \Delta x$$

This is illustrated in the leftmost figure above.

2. The trapezoidal rule approximates each subintegral by the area of a trapezoid with vertices at $(x_{j-1}, 0)$, $(x_{j-1}, f(x_{j-1}))$, $(x_j, f(x_j))$, $(x_j, 0)$:


$$\int_{x_{j-1}}^{x_j} f(x) dx \approx \frac{1}{2}[f(x_{j-1}) + f(x_j)] \Delta x$$

The trapezoid is illustrated in the middle figure above. We shall derive the formula for the area shortly.

3. Simpson's rule approximates two adjacent subintegrals by the area under a parabola that passes through the points $(x_{j-1}, f(x_{j-1}))$, $(x_j, f(x_j))$ and $(x_{j+1}, f(x_{j+1}))$:

$$\int_{x_{j-1}}^{x_{j+1}} f(x) dx \approx \frac{1}{3}[f(x_{j-1}) + 4f(x_j) + f(x_{j+1})] \Delta x$$

The parabola is illustrated in the right hand figure above. We shall derive the formula for the area shortly.

 Definition 1.11.1 Midpoints

In what follows we need to refer to the midpoint between x_{j-1} and x_j very frequently. To save on writing (and typing) we introduce the notation

$$\bar{x}_j = \frac{1}{2}(x_{j-1} + x_j).$$

The midpoint rule

The integral $\int_{x_{j-1}}^{x_j} f(x) dx$ represents the area between the curve $y = f(x)$ and the x -axis with x running from x_{j-1} to x_j . The width of this region is $x_j - x_{j-1} = \Delta x$. The height varies over the different values that $f(x)$ takes as x runs from x_{j-1} to x_j .

The midpoint rule approximates this area by the area of a rectangle of width $x_j - x_{j-1} = \Delta x$ and height $f(\bar{x}_j)$ which is the exact height at the midpoint of the range covered by x .



The area of the approximating rectangle is $f(\bar{x}_j)\Delta x$, and the midpoint rule approximates each subintegral by

$$\int_{x_{j-1}}^{x_j} f(x) dx \approx f(\bar{x}_j)\Delta x.$$

Applying this approximation to each subinterval and summing gives us the following approximation of the full integral:

$$\begin{aligned} \int_a^b f(x) dx &= \int_{x_0}^{x_1} f(x) dx + \int_{x_1}^{x_2} f(x) dx + \dots + \int_{x_{n-1}}^{x_n} f(x) dx \\ &\approx f(\bar{x}_1)\Delta x + f(\bar{x}_2)\Delta x + \dots + f(\bar{x}_n)\Delta x \end{aligned}$$

So notice that the approximation is the sum of the function evaluated at the midpoint of each interval and then multiplied by Δx . Our other approximations will have similar forms.

In summary:

 Equation 1.11.2 The midpoint rule

The midpoint rule approximation is

$$\int_a^b f(x) dx \approx [f(\bar{x}_1) + f(\bar{x}_2) + \dots + f(\bar{x}_n)] \Delta x$$

where $\Delta x = \frac{b-a}{n}$ and

$$\begin{array}{ccccccc} x_0 = a & x_1 = a + \Delta x & x_2 = a + 2\Delta x & \dots & x_{n-1} = b - \Delta x & x_n = b \\ \bar{x}_1 = \frac{x_0 + x_1}{2} & \bar{x}_2 = \frac{x_1 + x_2}{2} & \dots & \bar{x}_{n-1} = \frac{x_{n-2} + x_{n-1}}{2} & \bar{x}_n = \frac{x_{n-1} + x_n}{2} \end{array}$$

 Example 1.11.3 $\int_0^1 \frac{4}{1+x^2} dx$

We approximate the above integral using the midpoint rule with $n = 8$ step.

Solution

- First we set up all the x -values that we will need. Note that $a = 0$, $b = 1$, $\Delta x = \frac{1}{8}$ and

$$x_0 = 0 \quad x_1 = \frac{1}{8} \quad x_2 = \frac{2}{8} \quad \dots \quad x_7 = \frac{7}{8} \quad x_8 = \frac{8}{8} = 1$$

Consequently

$$\bar{x}_1 = \frac{1}{16} \quad \bar{x}_2 = \frac{3}{16} \quad \bar{x}_3 = \frac{5}{16} \quad \dots \quad \bar{x}_8 = \frac{15}{16}$$

- We now apply Equation 1.11.2 to the integrand $f(x) = \frac{4}{1+x^2}$:

$$\begin{aligned}
 \int_0^1 \frac{4}{1+x^2} dx &\approx \left[\overbrace{\frac{4}{1+\bar{x}_1^2}}^{f(\bar{x}_1)} + \overbrace{\frac{4}{1+\bar{x}_2^2}}^{f(\bar{x}_2)} + \cdots + \overbrace{\frac{4}{1+\bar{x}_7^2}}^{f(\bar{x}_{n-1})} + \overbrace{\frac{4}{1+\bar{x}_8^2}}^{f(\bar{x}_n)} \right] \Delta x \\
 &= \left[\frac{4}{1+\frac{1}{16^2}} + \frac{4}{1+\frac{3^2}{16^2}} + \frac{4}{1+\frac{5^2}{16^2}} + \frac{4}{1+\frac{7^2}{16^2}} + \frac{4}{1+\frac{9^2}{16^2}} \right. \\
 &\quad \left. + \frac{4}{1+\frac{11^2}{16^2}} + \frac{4}{1+\frac{13^2}{16^2}} + \frac{4}{1+\frac{15^2}{16^2}} \right] \frac{1}{8} \\
 &= [3.98444 + 3.86415 + 3.64413 + 3.35738 + 3.03858 + \\
 &\quad 2.71618 + 2.40941 + 2.12890] \frac{1}{8} \\
 &= 3.1429
 \end{aligned}$$

where we have rounded to four decimal places.

- In this case we can compute the integral exactly (which is one of the reasons it was chosen as a first example):

$$\int_0^1 \frac{4}{1+x^2} dx = 4 \arctan x \Big|_0^1 = \pi$$

- So the error in the approximation generated by eight steps of the midpoint rule is

$$|3.1429 - \pi| = 0.0013$$

- The relative error is then

$$\frac{|\text{approximate} - \text{exact}|}{\text{exact}} = \frac{|3.1429 - \pi|}{\pi} = 0.0004$$

That is the error is 0.0004 times the actual value of the integral.

- We can write this as a percentage error by multiplying it by 100

$$\text{percentage error} = 100 \times \frac{|\text{approximate} - \text{exact}|}{\text{exact}} = 0.04\%$$

That is, the error is about 0.04% of the exact value.

The midpoint rule gives us quite good estimates of the integral without too much work — though it is perhaps a little tedious to do by hand². Of course, it would be very helpful to quantify what we mean by “good” in this context and that requires us to discuss errors.

Definition 1.11.4

Suppose that α is an approximation to A . This approximation has

- absolute error $|A - \alpha|$ and
- relative error $\frac{|A - \alpha|}{A}$ and
- percentage error $100 \frac{|A - \alpha|}{A}$

We will discuss errors further in Section 1.11.4 below.

Example 1.11.5: $\int_0^\pi \sin x \, dx$

As a second example, we apply the midpoint rule with $n = 8$ steps to the above integral.

- We again start by setting up all the x -values that we will need. So $a = 0$, $b = \pi$, $\Delta x = \frac{\pi}{8}$ and

$$x_0 = 0 \quad x_1 = \frac{\pi}{8} \quad x_2 = \frac{2\pi}{8} \quad \cdots \quad x_7 = \frac{7\pi}{8} \quad x_8 = \frac{8\pi}{8} = \pi$$

Consequently,

$$\bar{x}_1 = \frac{\pi}{16} \quad \bar{x}_2 = \frac{3\pi}{16} \quad \dots \quad \bar{x}_7 = \frac{13\pi}{16} \quad \bar{x}_8 = \frac{15\pi}{16}$$

- Now apply Equation 1.11.2 to the integrand $f(x) = \sin x$:

$$\begin{aligned} \int_0^{\pi} \sin x \, dx &\approx \left[\sin(\bar{x}_1) + \sin(\bar{x}_2) + \dots + \sin(\bar{x}_8) \right] \Delta x \\ &= \left[\sin\left(\frac{\pi}{16}\right) + \sin\left(\frac{3\pi}{16}\right) + \sin\left(\frac{5\pi}{16}\right) + \sin\left(\frac{7\pi}{16}\right) + \sin\left(\frac{9\pi}{16}\right) + \right. \\ &\quad \left. \sin\left(\frac{11\pi}{16}\right) + \sin\left(\frac{13\pi}{16}\right) + \sin\left(\frac{15\pi}{16}\right) \right] \frac{\pi}{8} \\ &= \left[0.1951 + 0.5556 + 0.8315 + 0.9808 + 0.9808 + \right. \\ &\quad \left. 0.8315 + 0.5556 + 0.1951 \right] \times 0.3927 \\ &= 5.1260 \times 0.3927 = 2.013 \end{aligned}$$

- Again, we have chosen this example so that we can compare it against the exact value:

$$\int_0^{\pi} \sin x \, dx = [-\cos x]_0^{\pi} = -\cos \pi + \cos 0 = 2.$$

- So with eight steps of the midpoint rule we achieved

$$\begin{aligned} \text{absolute error} &= |2.013 - 2| = 0.013 \\ \text{relative error} &= \frac{|2.013 - 2|}{2} = 0.0065 \\ \text{percentage error} &= 100 \times \frac{|2.013 - 2|}{2} = 0.65\% \end{aligned}$$

With little work we have managed to estimate the integral to within 1% of its true value.

The trapezoidal rule

Consider again the area represented by the integral $\int_{x_{j-1}}^{x_j} f(x) \, dx$. The trapezoidal rule³ (unsurprisingly) approximates this area by a trapezoid⁴ whose vertices lie at

$$(x_{j-1}, 0), (x_{j-1}, f(x_{j-1})), (x_j, f(x_j)) \text{ and } (x_j, 0).$$



The trapezoidal approximation of the integral $\int_{x_{j-1}}^{x_j} f(x) \, dx$ is the shaded region in the figure on the right above. It has width $x_j - x_{j-1} = \Delta x$. Its left hand side has height $f(x_{j-1})$ and its right hand side has height $f(x_j)$.

As the figure below shows, the area of a trapezoid is its width times its average height.



So the trapezoidal rule approximates each subintegral by

$$\int_{x_{j-1}}^{x_j} f(x) \, dx \approx \frac{f(x_{j-1}) + f(x_j)}{2} \Delta x$$

Applying this approximation to each subinterval and then summing the result gives us the following approximation of the full integral

$$\begin{aligned} \int_a^b f(x) \, dx &= \int_{x_0}^{x_1} f(x) \, dx + \int_{x_1}^{x_2} f(x) \, dx + \dots + \int_{x_{n-1}}^{x_n} f(x) \, dx \\ &\approx \frac{f(x_0) + f(x_1)}{2} \Delta x + \frac{f(x_1) + f(x_2)}{2} \Delta x + \dots + \frac{f(x_{n-1}) + f(x_n)}{2} \Delta x \\ &= \left[\frac{1}{2} f(x_0) + f(x_1) + f(x_2) + \dots + f(x_{n-1}) + \frac{1}{2} f(x_n) \right] \Delta x \end{aligned}$$

So notice that the approximation has a very similar form to the midpoint rule, excepting that

- we evaluate the function at the x_j 's rather than at the midpoints, and

- we multiply the value of the function at the endpoints x_0, x_n by $\frac{1}{2}$.

In summary:

Equation 1.11.6 The trapezoidal rule

The trapezoidal rule approximation is

$$\int_a^b f(x) dx \approx \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \cdots + f(x_{n-1}) + \frac{1}{2}f(x_n) \right] \Delta x$$

where

$$\Delta x = \frac{b-a}{n}, \quad x_0 = a, \quad x_1 = a + \Delta x, \quad x_2 = a + 2\Delta x, \quad \cdots, \quad x_{n-1} = b - \Delta x, \quad x_n = b$$

To compare and contrast we apply the trapezoidal rule to the examples we did above with the midpoint rule.

✓ Example 1.11.7 $\int_0^1 \frac{4}{1+x^2} dx$ — using the trapezoidal rule

Solution

We proceed very similarly to Example 1.11.3 and again use $n = 8$ steps.

- We again have $f(x) = \frac{4}{1+x^2}$, $a = 0$, $b = 1$, $\Delta x = \frac{1}{8}$ and

$$x_0 = 0 \quad x_1 = \frac{1}{8} \quad x_2 = \frac{2}{8} \quad \cdots \quad x_7 = \frac{7}{8} \quad x_8 = \frac{8}{8} = 1$$

- Applying the trapezoidal rule, Equation 1.11.6, gives

$$\begin{aligned} \int_0^1 \frac{4}{1+x^2} dx &\approx \left[\frac{1}{2} \overbrace{\frac{4}{1+x_0^2}}^{f(x_0)} + \overbrace{\frac{4}{1+x_1^2}}^{f(x_1)} + \cdots + \overbrace{\frac{4}{1+x_7^2}}^{f(x_{n-1})} + \frac{1}{2} \overbrace{\frac{4}{1+x_8^2}}^{f(x_n)} \right] \Delta x \\ &= \left[\frac{1}{2} \frac{4}{1+0^2} + \frac{4}{1+\frac{1}{8^2}} + \frac{4}{1+\frac{2^2}{8^2}} + \frac{4}{1+\frac{3^2}{8^2}} \right. \\ &\quad \left. + \frac{4}{1+\frac{4^2}{8^2}} + \frac{4}{1+\frac{5^2}{8^2}} + \frac{4}{1+\frac{6^2}{8^2}} + \frac{4}{1+\frac{7^2}{8^2}} + \frac{1}{2} \frac{4}{1+\frac{8^2}{8^2}} \right] \frac{1}{8} \\ &= \left[\frac{1}{2} \times 4 + 3.939 + 3.765 + 3.507 \right. \\ &\quad \left. + 3.2 + 2.876 + 2.56 + 2.266 + \frac{1}{2} \times 2 \right] \frac{1}{8} \\ &= 3.139 \end{aligned}$$

to three decimal places.

- The exact value of the integral is still π . So the error in the approximation generated by eight steps of the trapezoidal rule is $|3.139 - \pi| = 0.0026$, which is $100 \frac{|3.139 - \pi|}{\pi} \% = 0.08\%$ of the exact answer. Notice that this is roughly twice the error that we achieved using the midpoint rule in Example 1.11.3.

Let us also redo Example 1.11.5 using the trapezoidal rule.

✓ Example 1.11.8 $\int_0^\pi \sin x dx$ — using the trapezoidal rule

Solution

We proceed very similarly to Example 1.11.5 and again use $n = 8$ steps.

- We again have $a = 0$, $b = \pi$, $\Delta x = \frac{\pi}{8}$ and

$$x_0 = 0 \quad x_1 = \frac{\pi}{8} \quad x_2 = \frac{2\pi}{8} \quad \cdots \quad x_7 = \frac{7\pi}{8} \quad x_8 = \frac{8\pi}{8} = \pi$$

- Applying the trapezoidal rule, Equation 1.11.6, gives

$$\begin{aligned} \int_0^\pi \sin x \, dx &\approx \left[\frac{1}{2} \sin(x_0) + \sin(x_1) + \cdots + \sin(x_7) + \frac{1}{2} \sin(x_8) \right] \Delta x \\ &= \left[\frac{1}{2} \sin 0 + \sin \frac{\pi}{8} + \sin \frac{2\pi}{8} + \sin \frac{3\pi}{8} + \sin \frac{4\pi}{8} + \sin \frac{5\pi}{8} \right. \\ &\quad \left. + \sin \frac{6\pi}{8} + \sin \frac{7\pi}{8} + \frac{1}{2} \sin \frac{8\pi}{8} \right] \frac{\pi}{8} \\ &= \left[\frac{1}{2} \times 0 + 0.3827 + 0.7071 + 0.9239 + 1.0000 + 0.9239 + \right. \\ &\quad \left. 0.7071 + 0.3827 + \frac{1}{2} \times 0 \right] \times 0.3927 \\ &= 5.0274 \times 0.3927 = 1.974 \end{aligned}$$

- The exact answer is $\int_0^\pi \sin x \, dx = -\cos x \Big|_0^\pi = 2$. So with eight steps of the trapezoidal rule we achieved $100 \frac{|1.974-2|}{2} = 1.3\%$ accuracy. Again this is approximately twice the error we achieved in Example 1.11.5 using the midpoint rule.

These two examples suggest that the midpoint rule is more accurate than the trapezoidal rule. Indeed, this observation is born out by a rigorous analysis of the error — see Section 1.11.4.

Simpson's Rule

When we use the trapezoidal rule we approximate the area $\int_{x_{j-1}}^{x_j} f(x) \, dx$ by the area between the x -axis and a straight line that runs from $(x_{j-1}, f(x_{j-1}))$ to $(x_j, f(x_j))$ — that is, we approximate the function $f(x)$ on this interval by a linear function that agrees with the function at each endpoint. An obvious way to extend this — just as we did when extending linear approximations to quadratic approximations in our differential calculus course — is to approximate the function with a quadratic. This is precisely what Simpson's⁵ rule does.

Simpson's rule approximates the integral over two neighbouring subintervals by the area between a parabola and the x -axis. In order to describe this parabola we need 3 distinct points (which is why we approximate two subintegrals at a time). That is, we approximate

$$\int_{x_0}^{x_1} f(x) \, dx + \int_{x_1}^{x_2} f(x) \, dx = \int_{x_0}^{x_2} f(x) \, dx$$

by the area bounded by the parabola that passes through the three points $(x_0, f(x_0))$, $(x_1, f(x_1))$ and $(x_2, f(x_2))$, the x -axis and the vertical lines $x = x_0$ and $x = x_2$.



We repeat this on the next pair of subintervals and approximate $\int_{x_2}^{x_4} f(x) \, dx$ by the area between the x -axis and the part of a parabola with $x_2 \leq x \leq x_4$. This parabola passes through the three points $(x_2, f(x_2))$, $(x_3, f(x_3))$ and $(x_4, f(x_4))$. And so on. Because Simpson's rule does the approximation two slices at a time, n must be even.

To derive Simpson's rule formula, we first find the equation of the parabola that passes through the three points $(x_0, f(x_0))$, $(x_1, f(x_1))$ and $(x_2, f(x_2))$. Then we find the area between the x -axis and the part of that parabola with $x_0 \leq x \leq x_2$. To simplify this computation consider a parabola passing through the points $(-h, y_{-1})$, $(0, y_0)$ and (h, y_1) .

Write the equation of the parabola as

$$y = Ax^2 + Bx + C$$

Then the area between it and the x -axis with x running from $-h$ to h is

$$\begin{aligned} \int_{-h}^h [Ax^2 + Bx + C] dx &= \left[\frac{A}{3}x^3 + \frac{B}{2}x^2 + Cx \right]_{-h}^h \\ &= \frac{2A}{3}h^3 + 2Ch && \text{it is helpful to write it as} \\ &= \frac{h}{3}(2Ah^2 + 6C) \end{aligned}$$

Now, the the three points $(-h, y_{-1})$, $(0, y_0)$ and (h, y_1) lie on this parabola if and only if

$$\begin{aligned} Ah^2 - Bh + C &= y_{-1} && \text{at } (-h, y_{-1}) \\ C &= y_0 && \text{at } (0, y_0) \\ Ah^2 + Bh + C &= y_1 && \text{at } (h, y_1) \end{aligned}$$

Adding the first and third equations together gives us

$$2Ah^2 + (B - B)h + 2C = y_{-1} + y_1$$

To this we add four times the middle equation

$$2Ah^2 + 6C = y_{-1} + 4y_0 + y_1.$$

This means that

$$\begin{aligned} \text{area} &= \int_{-h}^h [Ax^2 + Bx + C] dx = \frac{h}{3}(2Ah^2 + 6C) \\ &= \frac{h}{3}(y_{-1} + 4y_0 + y_1) \end{aligned}$$

Note that here

- h is one half of the length of the x -interval under consideration
- y_{-1} is the height of the parabola at the left hand end of the interval under consideration
- y_0 is the height of the parabola at the middle point of the interval under consideration
- y_1 is the height of the parabola at the right hand end of the interval under consideration

So Simpson's rule approximates

$$\int_{x_0}^{x_2} f(x) dx \approx \frac{1}{3} \Delta x [f(x_0) + 4f(x_1) + f(x_2)]$$

and

$$\int_{x_2}^{x_4} f(x) dx \approx \frac{1}{3} \Delta x [f(x_2) + 4f(x_3) + f(x_4)]$$

and so on. Summing these all together gives:

$$\begin{aligned} \int_a^b f(x) dx &= \int_{x_0}^{x_2} f(x) dx + \int_{x_2}^{x_4} f(x) dx + \int_{x_4}^{x_6} f(x) dx + \cdots + \int_{x_{n-2}}^{x_n} f(x) dx \\ &\approx \frac{\Delta x}{3} [f(x_0) + 4f(x_1) + f(x_2)] + \frac{\Delta x}{3} [f(x_2) + 4f(x_3) + f(x_4)] \\ &\quad + \frac{\Delta x}{3} [f(x_4) + 4f(x_5) + f(x_6)] + \cdots + \frac{\Delta x}{3} [f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)] \\ &= \left[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \cdots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n) \right] \frac{\Delta x}{3} \end{aligned}$$

In summary

Equation 1.11.9 Simpson's rule

The Simpson's rule approximation is

$$\int_a^b f(x) dx \approx \left[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \dots \right. \\ \left. \dots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n) \right] \frac{\Delta x}{3}$$

where n is even and

$$\Delta x = \frac{b-a}{n}, \quad x_0 = a, \quad x_1 = a + \Delta x, \quad x_2 = a + 2\Delta x, \quad \dots, \quad x_{n-1} = b - \Delta x, \quad x_n = b$$

Notice that Simpson's rule requires essentially no more work than the trapezoidal rule. In both rules we must evaluate $f(x)$ at $x = x_0, x_1, \dots, x_n$, but we add those terms multiplied by different constants ⁶.

Let's put it to work on our two running examples.

✓ Example 1.11.10 $\int_0^1 \frac{4}{1+x^2} dx$ — using Simpson's rule

Solution

We proceed almost identically to Example 1.11.7 and again use $n = 8$ steps.

- We have the same $\Delta, a, b, x_0, \dots, x_n$ as Example 1.11.7.
- Applying Equation 1.11.9 gives

$$\int_0^1 \frac{4}{1+x^2} dx \\ \approx \left[\frac{4}{1+0^2} + 4 \frac{4}{1+\frac{1}{8^2}} + 2 \frac{4}{1+\frac{2^2}{8^2}} + 4 \frac{4}{1+\frac{3^2}{8^2}} + 2 \frac{4}{1+\frac{4^2}{8^2}} \right. \\ \left. + 4 \frac{4}{1+\frac{5^2}{8^2}} + 2 \frac{4}{1+\frac{6^2}{8^2}} + 4 \frac{4}{1+\frac{7^2}{8^2}} + \frac{4}{1+\frac{8^2}{8^2}} \right] \frac{1}{8 \times 3} \\ = \left[4 + 4 \times 3.938461538 + 2 \times 3.764705882 + 4 \times 3.506849315 + 2 \times 3.2 \right. \\ \left. + 4 \times 2.876404494 + 2 \times 2.56 + 4 \times 2.265486726 + 2 \right] \frac{1}{8 \times 3} \\ = 3.14159250$$

to eight decimal places.

- This agrees with π (the exact value of the integral) to six decimal places. So the error in the approximation generated by eight steps of Simpson's rule is $|3.14159250 - \pi| = 1.5 \times 10^{-7}$, which is $100 \frac{|3.14159250 - \pi|}{\pi} \% = 5 \times 10^{-6} \%$ of the exact answer.

It is striking that the absolute error approximating with Simpson's rule is so much smaller than the error from the midpoint and trapezoidal rules.

$$\begin{aligned} \text{midpoint error} &= 0.0013 \\ \text{trapezoid error} &= 0.0026 \\ \text{Simpson error} &= 0.00000015 \end{aligned}$$

Buoyed by this success, we will also redo Example 1.11.8 using Simpson's rule.

✓ Example 1.11.11 $\int_0^\pi \sin x dx$ — Simpson's rule

Solution

We proceed almost identically to Example 1.11.8 and again use $n = 8$ steps.

- We have the same $\Delta, a, b, x_0, \dots, x_n$ as Example 1.11.7.
- Applying Equation 1.11.9 gives

$$\begin{aligned}
 \int_0^{\pi} \sin x \, dx & \approx \left[\sin(x_0) + 4 \sin(x_1) + 2 \sin(x_2) + \dots + 4 \sin(x_7) + \sin(x_8) \right] \frac{\Delta x}{3} \\
 & = \left[\sin(0) + 4 \sin\left(\frac{\pi}{8}\right) + 2 \sin\left(\frac{2\pi}{8}\right) + 4 \sin\left(\frac{3\pi}{8}\right) + 2 \sin\left(\frac{4\pi}{8}\right) \right. \\
 & \quad \left. + 4 \sin\left(\frac{5\pi}{8}\right) + 2 \sin\left(\frac{6\pi}{8}\right) + 4 \sin\left(\frac{7\pi}{8}\right) + \sin\left(\frac{8\pi}{8}\right) \right] \frac{\pi}{8 \times 3} \\
 & = \left[0 + 4 \times 0.382683 + 2 \times 0.707107 + 4 \times 0.923880 + 2 \times 1.0 \right. \\
 & \quad \left. + 4 \times 0.923880 + 2 \times 0.707107 + 4 \times 0.382683 + 0 \right] \frac{\pi}{8 \times 3} \\
 & = 15.280932 \times 0.130900 \\
 & = 2.00027
 \end{aligned}$$

- With only eight steps of Simpson's rule we achieved $100 \frac{2.00027-2}{2} = 0.014\%$ accuracy.

Again we contrast the error we achieved with the other two rules:

$$\begin{aligned}
 \text{midpoint error} & = 0.013 \\
 \text{trapezoid error} & = 0.026 \\
 \text{Simpson error} & = 0.00027
 \end{aligned}$$

This completes our derivation of the midpoint, trapezoidal and Simpson's rules for approximating the values of definite integrals. So far we have not attempted to see how efficient and how accurate the algorithms are in general. That's our next task.

Three Simple Numerical Integrators — Error Behaviour

Now we are armed with our three (relatively simple) method for numerical integration we should give thought to how practical they might be in the real world⁷. Two obvious considerations when deciding whether or not a given algorithm is of any practical value are

1. the amount of computational effort required to execute the algorithm and
2. the accuracy that this computational effort yields.

For algorithms like our simple integrators, the bulk of the computational effort usually goes into evaluating the function $f(x)$. The number of evaluations of $f(x)$ required for n steps of the midpoint rule is n , while the number required for n steps of the trapezoidal and Simpson's rules is $n + 1$. So all three of our rules require essentially the same amount of effort — one evaluation of $f(x)$ per step.

To get a first impression of the error behaviour of these methods, we apply them to a problem whose answer we know exactly:

$$\int_0^{\pi} \sin x \, dx = -\cos x \Big|_0^{\pi} = 2.$$

To be a little more precise, we would like to understand how the errors of the three methods change as we increase the effort we put in (as measured by the number of steps n). The following table lists the error in the approximate value for this number generated by our three rules applied with three different choices of n . It also lists the number of evaluations of f required to compute the approximation.

n	<i>Midpoint</i>		<i>Trapezoidal</i>		<i>Simpson's</i>	
	error	# evals	error	# evals	error	# evals
10	8.2×10^{-3}	10	1.6×10^{-2}	11	1.1×10^{-4}	11
100	8.2×10^{-5}	100	1.6×10^{-4}	101	1.1×10^{-8}	101
1000	8.2×10^{-7}	1000	1.6×10^{-6}	1001	1.1×10^{-12}	1001

Observe that

- Using 101 evaluations of f worth of Simpson's rule gives an error 75 times smaller than 1000 evaluations of f worth of the midpoint rule.
- The trapezoidal rule error with n steps is about twice the midpoint rule error with n steps.
- With the midpoint rule, increasing the number of steps by a factor of 10 appears to reduce the error by about a factor of $100 = 10^2 = n^2$.
- With the trapezoidal rule, increasing the number of steps by a factor of 10 appears to reduce the error by about a factor of $10^2 = n^2$.
- With Simpson's rule, increasing the number of steps by a factor of 10 appears to reduce the error by about a factor of $10^4 = n^4$.

So it looks like

$$\begin{aligned} \text{approx value of } \int_a^b f(x) dx \text{ given by } n \text{ midpoint steps} &\approx \int_a^b f(x) dx + K_M \cdot \frac{1}{n^2} \\ \text{approx value of } \int_a^b f(x) dx \text{ given by } n \text{ trapezoidal steps} &\approx \int_a^b f(x) dx + K_T \cdot \frac{1}{n^2} \\ \text{approx value of } \int_a^b f(x) dx \text{ given by } n \text{ Simpson's steps} &\approx \int_a^b f(x) dx + K_S \cdot \frac{1}{n^4} \end{aligned}$$

with some constants K_M , K_T and K_S . It also seems that $K_T \approx 2K_M$.



Figure 1.11.12. A log-log plot of the error in the n step approximation to $\int_0^\pi \sin x dx$.

To test these conjectures for the behaviour of the errors we apply our three rules with about ten different choices of n of the form $n = 2^m$ with m integer. Figure 1.11.12 contains two graphs of the results. The left-hand plot shows the results for the midpoint and trapezoidal rules and the right-hand plot shows the results for Simpson's rule.

For each rule we are expecting (based on our conjectures above) that the error

$$e_n = |\text{exact value} - \text{approximate value}|$$

with n steps is (roughly) of the form

$$e_n = K \frac{1}{n^k}$$

for some constants K and k . We would like to test if this is really the case, by graphing $Y = e_n$ against $X = n$ and seeing if the graph “looks right”. But it is not easy to tell whether or not a given curve really is $Y = \frac{K}{X^k}$, for some specific k , by just looking at it. However, your eye is pretty good at determining whether or not a graph is a straight line. Fortunately, there is a little trick that turns the curve $Y = \frac{K}{X^k}$ into a straight line — no matter what k is.

Instead of plotting Y against X , we plot $\log Y$ against $\log X$. This transformation⁸ works because when $Y = \frac{K}{X^k}$

$$\log Y = \log K - k \log X$$

So plotting $y = \log Y$ against $x = \log X$ gives the straight line $y = \log K - kx$, which has slope $-k$ and y -intercept $\log K$.

The three graphs in Figure 1.11.12 plot $y = \log_2 e_n$ against $x = \log_2 n$ for our three rules. Note that we have chosen to use logarithms⁹ with this “unusual base” because it makes it very clear how much the error is improved if we *double* the number of steps used. To be more precise — one unit step along the x -axis represents changing $n \mapsto 2n$. For example, applying Simpson's rule with $n = 2^4$ steps results in an error of 0000166, so the point $(x = \log_2 2^4 = 4, y = \log_2 0000166 = \frac{\log 0000166}{\log 2} = -15.8)$ has been included on the graph. Doubling the effort used — that is, doubling the number of steps to $n = 2^5$ — results in an error of 0.00000103. So, the data point $(x = \log_2 2^5 = 5, y = \log_2 0.00000103 = \frac{\ln 0.00000103}{\ln 2} = -19.9)$ lies on the graph. Note that the x -coordinates of these points differ by 1 unit.

For each of the three sets of data points, a straight line has also been plotted “through” the data points. A procedure called linear regression¹⁰ has been used to decide precisely which straight line to plot. It provides a formula for the slope and y -intercept of the straight line which “best fits” any given set of data points. From the three lines, it sure looks like $k = 2$ for the midpoint and

trapezoidal rules and $k = 4$ for Simpson's rule. It also looks like the ratio between the value of K for the trapezoidal rule, namely $K = 2^{0.7253}$, and the value of K for the midpoint rule, namely $K = 2^{-0.2706}$, is pretty close to 2: $2^{0.7253}/2^{-0.2706} = 2^{0.9959}$.

The intuition, about the error behaviour, that we have just developed is in fact correct — provided the integrand $f(x)$ is reasonably smooth. To be more precise

Theorem 1.11.13 Numerical integration errors

Assume that $|f''(x)| \leq M$ for all $a \leq x \leq b$. Then

the total error introduced by the midpoint rule is bounded by $\frac{M}{24} \frac{(b-a)^3}{n^2}$

and

the total error introduced by the trapezoidal rule is bounded by $\frac{M}{12} \frac{(b-a)^3}{n^2}$

when approximating $\int_a^b f(x) dx$. Further, if $|f^{(4)}(x)| \leq L$ for all $a \leq x \leq b$, then

the total error introduced by Simpson's rule is bounded by $\frac{L}{180} \frac{(b-a)^5}{n^4}$.

The first of these error bounds is proven in the following (optional) section. Here are some examples which illustrate how they are used. First let us check that the above result is consistent with our data in Figure 1.11.12

Example 1.11.14 Midpoint rule error approximating $\int_0^\pi \sin x dx$

- The integral $\int_0^\pi \sin x dx$ has $b - a = \pi$.
- The second derivative of the integrand satisfies

$$\left| \frac{d^2}{dx^2} \sin x \right| = |-\sin x| \leq 1$$

So we take $M = 1$.

- So the error, e_n , introduced when n steps are used is bounded by

$$\begin{aligned} |e_n| &\leq \frac{M}{24} \frac{(b-a)^3}{n^2} \\ &= \frac{\pi^3}{24} \frac{1}{n^2} \\ &\approx 1.29 \frac{1}{n^2} \end{aligned}$$

- The data in the graph in Figure 1.11.12 gives

$$|e_n| \approx 2^{-0.2706} \frac{1}{n^2} = 0.83 \frac{1}{n^2}$$

which is consistent with the bound $|e_n| \leq \frac{\pi^3}{24} \frac{1}{n^2}$.

In a typical application we would be asked to evaluate a given integral to some specified accuracy. For example, if you are manufacturer and your machinery can only cut materials to an accuracy of $\frac{1}{10}$ th of a millimeter, there is no point in making design specifications more accurate than $\frac{1}{10}$ th of a millimeter.

✓ Example 1.11.15 How many steps for a given accuracy?

Suppose, for example, that we wish to use the midpoint rule to evaluate¹¹

$$\int_0^1 e^{-x^2} dx$$

to within an accuracy of 10^{-6} .

Solution

- The integral has $a = 0$ and $b = 1$.
- The first two derivatives of the integrand are

$$\begin{aligned} \frac{d}{dx} e^{-x^2} &= -2xe^{-x^2} && \text{and} \\ \frac{d^2}{dx^2} e^{-x^2} &= \frac{d}{dx} (-2xe^{-x^2}) = -2e^{-x^2} + 4x^2 e^{-x^2} = 2(2x^2 - 1)e^{-x^2} \end{aligned}$$

- As x runs from 0 to 1, $2x^2 - 1$ increases from -1 to 1 , so that

$$0 \leq x \leq 1 \implies |2x^2 - 1| \leq 1, e^{-x^2} \leq 1 \implies |2(2x^2 - 1)e^{-x^2}| \leq 2$$

So we take $M = 2$.

- The error introduced by the n step midpoint rule is at most

$$\begin{aligned} e_n &\leq \frac{M}{24} \frac{(b-a)^3}{n^2} \\ &\leq \frac{2}{24} \frac{(1-0)^3}{n^2} = \frac{1}{12n^2} \end{aligned}$$

- We need this error to be smaller than 10^{-6} so

$$\begin{aligned} e_n &\leq \frac{1}{12n^2} \leq 10^{-6} && \text{and so} \\ 12n^2 &\geq 10^6 && \text{clean up} \\ n^2 &\geq \frac{10^6}{12} = 83333.3 && \text{square root both sides} \\ n &\geq 288.7 \end{aligned}$$

So 289 steps of the midpoint rule will do the job.

- In fact $n = 289$ results in an error of about 3.7×10^{-7} .

That seems like far too much work, and the trapezoidal rule will have twice the error. So we should look at Simpson's rule.

✓ Example 1.11.16 How many steps using Simpson's rule?

Suppose now that we wish evaluate $\int_0^1 e^{-x^2} dx$ to within an accuracy of 10^{-6} — but now using Simpson's rule. How many steps should we use?

Solution

- Again we have $a = 0$, $b = 1$.
- We then need to bound $\frac{d^4}{dx^4} e^{-x^2}$ on the domain of integration, $0 \leq x \leq 1$.

$$\begin{aligned}\frac{d^3}{dx^3}e^{-x^2} &= \frac{d}{dx} \{2(2x^2 - 1)e^{-x^2}\} = 8xe^{-x^2} - 4x(2x^2 - 1)e^{-x^2} \\ &= 4(-2x^3 + 3x)e^{-x^2} \\ \frac{d^4}{dx^4}e^{-x^2} &= \frac{d}{dx} \{4(-2x^3 + 3x)e^{-x^2}\} \\ &= 4(-6x^2 + 3)e^{-x^2} - 8x(-2x^3 + 3x)e^{-x^2} \\ &= 4(4x^4 - 12x^2 + 3)e^{-x^2}\end{aligned}$$

- Now, for any x , $e^{-x^2} \leq 1$. Also, for $0 \leq x \leq 1$,

$$\begin{aligned}0 &\leq x^2, x^4 \leq 1 && \text{so} \\ 3 &\leq 4x^4 + 3 \leq 7 && \text{and} \\ -12 &\leq -12x^2 \leq 0 && \text{adding these together gives} \\ -9 &\leq 4x^4 - 12x^2 + 3 \leq 7\end{aligned}$$

Consequently, $|4x^4 - 12x^2 + 3|$ is bounded by 9 and so

$$\left| \frac{d^4}{dx^4}e^{-x^2} \right| \leq 4 \times 9 = 36$$

So take $L = 36$.

- The error introduced by the n step Simpson's rule is at most

$$\begin{aligned}e_n &\leq \frac{L}{180} \frac{(b-a)^5}{n^4} \\ &\leq \frac{36}{180} \frac{(1-0)^5}{n^4} = \frac{1}{5n^4}\end{aligned}$$

- In order for this error to be no more than 10^{-6} we require n to satisfy

$$\begin{aligned}e_n &\leq \frac{1}{5n^4} \leq 10^{-6} && \text{and so} \\ 5n^4 &\geq 10^6 \\ n^4 &\geq 200000 && \text{take fourth root} \\ n &\geq 21.15\end{aligned}$$

So 22 steps of Simpson's rule will do the job.

- $n = 22$ steps actually results in an error of 3.5×10^{-8} . The reason that we get an error so much smaller than we need is that we have overestimated the number of steps required. This, in turn, occurred because we made quite a rough bound of $\left| \frac{d^4}{dx^4}f(x) \right| \leq 36$. If we are more careful then we will get a slightly smaller n . It actually turns out¹² that you only need $n = 10$ to approximate within 10^{-6} .

Optional — An error bound for the midpoint rule

We now try develop some understanding as to why we got the above experimental results. We start with the error generated by a single step of the midpoint rule. That is, the error introduced by the approximation

$$\int_{x_0}^{x_1} f(x) dx \approx f(\bar{x}_1)\Delta x \quad \text{where } \Delta x = x_1 - x_0, \bar{x}_1 = \frac{x_0 + x_1}{2}$$

To do this we are going to need to apply integration by parts in a sneaky way. Let us start by considering¹³ a subinterval $\alpha \leq x \leq \beta$ and let's call the width of the subinterval $2q$ so that $\beta = \alpha + 2q$. If we were to now apply the midpoint rule to this subinterval, then we would write

$$\int_{\alpha}^{\beta} f(x) dx \approx 2q \cdot f(\alpha + q) = qf(\alpha + q) + qf(\beta - q)$$

since the interval has width $2q$ and the midpoint is $\alpha + q = \beta - q$.

The sneaky trick we will employ is to write

$$\int_{\alpha}^{\beta} f(x) dx = \int_{\alpha}^{\alpha+q} f(x) dx + \int_{\beta-q}^{\beta} f(x) dx$$

and then examine each of the integrals on the right-hand side (using integration by parts) and show that they are each of the form

$$\int_{\alpha}^{\alpha+q} f(x) dx \approx qf(\alpha+q) + \text{small error term}$$

$$\int_{\beta-q}^{\beta} f(x) dx \approx qf(\beta-q) + \text{small error term}$$

Let us apply integration by parts to $\int_{\alpha}^{\alpha+q} f(x) dx$ — with $u = f(x)$, $dv = dx$ so $du = f'(x) dx$ and we will make the slightly non-standard choice of $v = x - \alpha$:

$$\begin{aligned} \int_{\alpha}^{\alpha+q} f(x) dx &= [(x - \alpha)f(x)]_{\alpha}^{\alpha+q} - \int_{\alpha}^{\alpha+q} (x - \alpha)f'(x) dx \\ &= qf(\alpha+q) - \int_{\alpha}^{\alpha+q} (x - \alpha)f'(x) dx \end{aligned}$$

Notice that the first term on the right-hand side is the term we need, and that our non-standard choice of v allowed us to avoid introducing an $f(\alpha)$ term.

Now integrate by parts again using $u = f'(x)$, $dv = (x - \alpha) dx$, so $du = f''(x)$, $v = \frac{(x-\alpha)^2}{2}$:

$$\begin{aligned} \int_{\alpha}^{\alpha+q} f(x) dx &= qf(\alpha+q) - \int_{\alpha}^{\alpha+q} (x - \alpha)f'(x) dx \\ &= qf(\alpha+q) - \left[\frac{(x - \alpha)^2}{2} f'(x) \right]_{\alpha}^{\alpha+q} + \int_{\alpha}^{\alpha+q} \frac{(x - \alpha)^2}{2} f''(x) dx \\ &= qf(\alpha+q) - \frac{q^2}{2} f'(\alpha+q) + \int_{\alpha}^{\alpha+q} \frac{(x - \alpha)^2}{2} f''(x) dx \end{aligned}$$

To obtain a similar expression for the other integral, we repeat the above steps and obtain:

$$\int_{\beta-q}^{\beta} f(x) dx = qf(\beta-q) + \frac{q^2}{2} f'(\beta-q) + \int_{\beta-q}^{\beta} \frac{(x - \beta)^2}{2} f''(x) dx$$

Now add together these two expressions

$$\begin{aligned} \int_{\alpha}^{\alpha+q} f(x) dx + \int_{\beta-q}^{\beta} f(x) dx &= qf(\alpha+q) + qf(\beta-q) + \frac{q^2}{2} (f'(\beta-q) - f'(\alpha+q)) \\ &\quad + \int_{\alpha}^{\alpha+q} \frac{(x - \alpha)^2}{2} f''(x) dx + \int_{\beta-q}^{\beta} \frac{(x - \beta)^2}{2} f''(x) dx \end{aligned}$$

Then since $\alpha + q = \beta - q$ we can combine the integrals on the left-hand side and eliminate some terms from the right-hand side:

$$\int_{\alpha}^{\beta} f(x) dx = 2qf(\alpha+q) + \int_{\alpha}^{\alpha+q} \frac{(x - \alpha)^2}{2} f''(x) dx + \int_{\beta-q}^{\beta} \frac{(x - \beta)^2}{2} f''(x) dx$$

Rearrange this expression a little and take absolute values

$$\left| \int_{\alpha}^{\beta} f(x) dx - 2qf(\alpha+q) \right| \leq \left| \int_{\alpha}^{\alpha+q} \frac{(x - \alpha)^2}{2} f''(x) dx \right| + \left| \int_{\beta-q}^{\beta} \frac{(x - \beta)^2}{2} f''(x) dx \right|$$

where we have also made use of the triangle inequality¹⁴. By assumption $|f''(x)| \leq M$ on the interval $\alpha \leq x \leq \beta$, so

$$\begin{aligned} \left| \int_{\alpha}^{\beta} f(x) dx - 2qf(\alpha + q) \right| &\leq M \int_{\alpha}^{\alpha+q} \frac{(x-\alpha)^2}{2} dx + M \int_{\beta-q}^{\beta} \frac{(x-\beta)^2}{2} dx \\ &= \frac{Mq^3}{3} = \frac{M(\beta-\alpha)^3}{24} \end{aligned}$$

where we have used $q = \frac{\beta-\alpha}{2}$ in the last step.

Thus on any interval $x_i \leq x \leq x_{i+1} = x_i + \Delta x$

$$\left| \int_{x_i}^{x_{i+1}} f(x) dx - \Delta x f\left(\frac{x_i + x_{i+1}}{2}\right) \right| \leq \frac{M}{24} (\Delta x)^3$$

Putting everything together we see that the error using the midpoint rule is bounded by

$$\begin{aligned} &\left| \int_a^b f(x) dx - [f(\bar{x}_1) + f(\bar{x}_2) + \cdots + f(\bar{x}_n)] \Delta x \right| \\ &\leq \left| \int_{x_0}^{x_1} f(x) dx - \Delta x f(\bar{x}_1) \right| + \cdots + \left| \int_{x_{n-1}}^{x_n} f(x) dx - \Delta x f(\bar{x}_n) \right| \\ &\leq n \times \frac{M}{24} (\Delta x)^3 = n \times \frac{M}{24} \left(\frac{b-a}{n}\right)^3 = \frac{M(b-a)^3}{24n^2} \end{aligned}$$

as required.

A very similar analysis shows that, as was stated in Theorem 1.11.13 above,

- the total error introduced by the trapezoidal rule is bounded by $\frac{M}{12} \frac{(b-a)^3}{n^2}$,
- the total error introduced by Simpson's rule is bounded by $\frac{M}{180} \frac{(b-a)^5}{n^4}$

Exercises

Recall that we are using $\log x$ to denote the logarithm of x with base e . In other courses it is often denoted $\ln x$.

Stage 1

? 1

Suppose we approximate an object to have volume 1.5m^3 , when its exact volume is 1.387m^3 . Give the relative error, absolute error, and percent error of our approximation.

? 2

Consider approximating $\int_2^{10} f(x) dx$, where $f(x)$ is the function in the graph below.



1. Draw the rectangles associated with the midpoint rule approximation and $n = 4$.
2. Draw the trapezoids associated with the trapezoidal rule approximation and $n = 4$.

You don't have to give an approximation.

? 3

Let $f(x) = -\frac{1}{12}x^4 + \frac{7}{6}x^3 - 3x^2$.

1. Find a reasonable value M such that $|f''(x)| \leq M$ for all $1 \leq x \leq 6$.
2. Find a reasonable value L such that $|f^{(4)}(x)| \leq L$ for all $1 \leq x \leq 6$.

? 4

Let $f(x) = x \sin x + 2 \cos x$. Find a reasonable value M such that $|f''(x)| \leq M$ for all $-3 \leq x \leq 2$.

? 5

Consider the quantity $A = \int_{-\pi}^{\pi} \cos x \, dx$.

1. Find the upper bound on the error using Simpson's rule with $n = 4$ to approximate A using Theorem 1.11.13 in the text.
2. Find the Simpson's rule approximation of A using $n = 4$.
3. What is the (actual) absolute error in the Simpson's rule approximation of A with $n = 4$?

? 6

Give a function $f(x)$ such that:

- $f''(x) \leq 3$ for every x in $[0, 1]$, and
- the error using the trapezoidal rule approximating $\int_0^1 f(x) \, dx$ with $n = 2$ intervals is exactly $\frac{1}{16}$.

? 7

Suppose my mother is under 100 years old, and I am under 200 years old. ¹⁵We're going somewhere with this. Who is older?

? 8

1. True or False: for fixed positive constants M, n, a , and b , with $b > a$,

$$\frac{M}{24} \frac{(b-a)^3}{n^2} \leq \frac{M}{12} \frac{(b-a)^3}{n^2}$$

2. True or False: for a function $f(x)$ and fixed constants n, a , and b , with $b > a$, the n -interval midpoint approximation of $\int_a^b f(x) \, dx$ is more accurate than the n -interval trapezoidal approximation.

? 9 (*)

Decide whether the following statement is true or false. If false, provide a counterexample. If true, provide a brief justification.

When $f(x)$ is positive and concave up, any trapezoidal rule approximation for $\int_a^b f(x) \, dx$ will be an upper estimate for $\int_a^b f(x) \, dx$.

? 10

Give a polynomial $f(x)$ with the property that the Simpson's rule approximation of $\int_a^b f(x) \, dx$ is exact for all a, b , and n .

Stage 2

Questions 11 and 12 ask you to approximate a given integral using the formulas in Equations 1.11.2, 1.11.6, and 1.11.9 in the text.

Questions 13 through 17 ask you to approximate a quantity based on observed data.

In Questions 18 through 24, we practice finding error bounds for our approximations.

? 11

Write out all three approximations of $\int_0^{30} \frac{1}{x^3+1} dx$ with $n = 6$. (That is: midpoint, trapezoidal, and Simpson's.) You do not need to simplify your answers.

? 12 (*)

Find the midpoint rule approximation to $\int_0^\pi \sin x dx$ with $n = 3$.

? 13 (*)

The solid V is 40 cm high and the horizontal cross sections are circular disks. The table below gives the diameters of the cross sections in centimeters at 10 cm intervals. Use the trapezoidal rule to estimate the volume of V .

height	0	10	20	30	40
diameter	24	16	10	6	4

? 14 (*)

A 6 metre long cedar log has cross sections that are approximately circular. The diameters of the log, measured at one metre intervals, are given below:

metres from left end of log	0	1	2	3	4	5	6
diameter in metres	1.2	1	0.8	0.8	1	1	1.2

Use Simpson's Rule to estimate the volume of the log.

? 15 (*)

The circumference of an 8 metre high tree at different heights above the ground is given in the table below. Assume that all horizontal cross-sections of the tree are circular disks.

height (metres)	0	2	4	6	8
circumference (metres)	1.2	1.1	1.3	0.9	0.2

Use Simpson's rule to approximate the volume of the tree.

? 16 (*)

By measuring the areas enclosed by contours on a topographic map, a geologist determines the cross sectional areas A in m^2 of a 60 m high hill. The table below gives the cross sectional area $A(h)$ at various heights h . The volume of the hill is $V = \int_0^{60} A(h) dh$.

h	0	10	20	30	40	50	60
A	10,200	9,200	8,000	7,100	4,500	2,400	100

1. If the geologist uses the Trapezoidal Rule to estimate the volume of the hill, what will be their estimate, to the nearest 1,000 m³?
2. What will be the geologist's estimate of the volume of the hill if they use Simpson's Rule instead of the Trapezoidal Rule?

? 17 (*)

The graph below applies to both parts (a) and (b).



1. Use the Trapezoidal Rule, with $n = 4$, to estimate the area under the graph between $x = 2$ and $x = 6$. Simplify your answer completely.
2. Use Simpson's Rule, with $n = 4$, to estimate the area under the graph between $x = 2$ and $x = 6$.

? 18 (*)

The integral $\int_{-1}^1 \sin(x^2) dx$ is estimated using the Midpoint Rule with 1000 intervals. Show that the absolute error in this approximation is at most $2 \cdot 10^{-6}$.

You may use the fact that when approximating $\int_a^b f(x) dx$ with the Midpoint Rule using n points, the absolute value of the error is at most $M(b-a)^3/24n^2$ when $|f''(x)| \leq M$ for all $x \in [a, b]$.

? 19 (*)

The total error using the midpoint rule with n subintervals to approximate the integral of $f(x)$ over $[a, b]$ is bounded by $\frac{M(b-a)^3}{(24n^2)}$, if $|f''(x)| \leq M$ for all $a \leq x \leq b$.

Using this bound, if the integral $\int_{-2}^1 2x^4 dx$ is approximated using the midpoint rule with 60 subintervals, what is the largest possible error between the approximation M_{60} and the true value of the integral?

? 20 (*)

Both parts of this question concern the integral $I = \int_0^2 (x-3)^5 dx$.

1. Write down the Simpson's Rule approximation to I with $n = 6$. Leave your answer in calculator-ready form.
2. Which method of approximating I results in a smaller error bound: the Midpoint Rule with $n = 100$ intervals, or Simpson's Rule with $n = 10$ intervals? You may use the formulas

$$|E_M| \leq \frac{M(b-a)^3}{24n^2} \quad \text{and} \quad |E_S| \leq \frac{L(b-a)^5}{180n^4},$$

where M is an upper bound for $|f''(x)|$ and L is an upper bound for $|f^{(4)}(x)|$, and E_M and E_S are the absolute errors arising from the midpoint rule and Simpson's rule, respectively.

? 21 (*)

Find a bound for the error in approximating $\int_1^5 \frac{1}{x} dx$ using Simpson's rule with $n = 4$. Do not write down the Simpson's rule approximation S_4 .

In general the error in approximating $\int_a^b f(x) dx$ using Simpson's rule with n steps is bounded by $\frac{L(b-a)}{180}(\Delta x)^4$ where $\Delta x = \frac{b-a}{n}$ and $L \geq |f^{(4)}(x)|$ for all $a \leq x \leq b$.

? 22 (*)

Find a bound for the error in approximating

$$\int_0^1 (e^{-2x} + 3x^3) dx$$

using Simpson's rule with $n = 6$. Do not write down the Simpson's rule approximation S_n .

In general, the error in approximating $\int_a^b f(x) dx$ using Simpson's rule with n steps is bounded by $\frac{L(b-a)}{180}(\Delta x)^4$ where $\Delta x = \frac{b-a}{n}$ and $L \geq |f^{(4)}(x)|$ for all $a \leq x \leq b$.

? 23 (*)

Let $I = \int_1^2 (1/x) dx$.

1. Write down the trapezoidal approximation T_4 for I . *You do not need to simplify your answer.*
2. Write down the Simpson's approximation S_4 for I . *You do not need to simplify your answer.*
3. Without computing I , find an upper bound for $|I - S_4|$. You may use the fact that if $|f^{(4)}(x)| \leq L$ on the interval $[a, b]$, then the error in using S_n to approximate $\int_a^b f(x) dx$ has absolute value less than or equal to $L(b-a)^5/180n^4$.

? 24 (*)

A function $s(x)$ satisfies $s(0) = 1.00664$, $s(2) = 1.00543$, $s(4) = 1.00435$, $s(6) = 1.00331$, $s(8) = 1.00233$. Also, it is known to satisfy $|s^{(k)}(x)| \leq \frac{k}{1000}$ for $0 \leq x \leq 8$ and all positive integers k .

1. Find the best Trapezoidal Rule and Simpson's Rule approximations that you can for $I = \int_0^8 s(x) dx$.
2. Determine the maximum possible sizes of errors in the approximations you gave in part (a). Recall that if a function $f(x)$ satisfies $|f^{(k)}(x)| \leq K_k$ on $[a, b]$, then

$$\left| \int_a^b f(x) dx - T_n \right| \leq \frac{K_2(b-a)^3}{12n^2} \quad \text{and} \quad \left| \int_a^b f(x) dx - S_n \right| \leq \frac{K_4(b-a)^5}{180n^4}$$

? 25 (*)

Consider the trapezoidal rule for making numerical approximations to $\int_a^b f(x) dx$. The error for the trapezoidal rule satisfies $|E_T| \leq \frac{M(b-a)^3}{12n^2}$, where $|f''(x)| \leq M$ for $a \leq x \leq b$. If $-2 < f''(x) < 0$ for $1 \leq x \leq 4$, find a value of n to guarantee the trapezoidal rule will give an approximation for $\int_1^4 f(x) dx$ with absolute error, $|E_T|$, less than 0.001.

Stage 3

? 26 (*)

A swimming pool has the shape shown in the figure below. The vertical cross-sections of the pool are semi-circular disks. The distances in feet across the pool are given in the figure at 2-foot intervals along the sixteen-foot length of the pool. Use Simpson's Rule to estimate the volume of the pool.



? 27 (*)

A piece of wire 1m long with radius 1mm is made in such a way that the density varies in its cross-section, but is radially symmetric (that is, the local density $g(r)$ in kg/m^3 depends only on the distance r in mm from the center of the wire). Take as given that the total mass W of the wire in kg is given by

$$W = 2\pi 10^{-6} \int_0^1 r g(r) dr$$

Data from the manufacturer is given below:

r	0	1/4	1/2	3/4	1
$g(r)$	8051	8100	8144	8170	8190

1. Find the best Trapezoidal Rule approximation that you can for W based on the data in the table.
2. Suppose that it is known that $|g'(r)| < 200$ and $|g''(r)| < 150$ for all values of r . Determine the maximum possible size of the error in the approximation you gave in part (a). Recall that if a function $f(x)$ satisfies $|f''(x)| \leq M$ on $[a, b]$, then

$$|I - T_n| \leq \frac{M(b-a)^3}{12n^2}$$

where $I = \int_a^b f(x) dx$ and T_n is the Trapezoidal Rule approximation to I using n subintervals.

? 28 (*)

Simpson's rule can be used to approximate $\log 2$, since $\log 2 = \int_1^2 \frac{1}{x} dx$.

1. Use Simpson's rule with 6 subintervals to approximate $\log 2$.
2. How many subintervals are required in order to guarantee that the absolute error is less than 0.00001?

Note that if E_n is the error using n subintervals, then $|E_n| \leq \frac{L(b-a)^5}{180n^4}$ where L is the maximum absolute value of the fourth derivative of the function being integrated and a and b are the end points of the interval.

? 29 (*)

Let $I = \int_0^2 \cos(x^2) dx$ and let S_n be the Simpson's rule approximation to I using n subintervals.

1. Estimate the maximum absolute error in using S_8 to approximate I .
2. How large should n be in order to ensure that $|I - S_n| \leq 0.0001$?

Note: The graph of $f''''(x)$, where $f(x) = \cos(x^2)$, is shown below. The absolute error in the Simpson's rule approximation is bounded by $\frac{L(b-a)^5}{180n^4}$ when $|f''''(x)| \leq L$ on the interval $[a, b]$.



? 30 (*)

Define a function $f(x)$ and an integral I by

$$f(x) = \int_0^{x^2} \sin(\sqrt{t}) dt, \quad I = \int_0^1 f(t) dt$$

Estimate how many subdivisions are needed to calculate I to five decimal places of accuracy using the trapezoidal rule.

Note that if E_n is the error using n subintervals, then $|E_n| \leq \frac{M(b-a)^3}{12n^2}$, where M is the maximum absolute value of the second derivative of the function being integrated and a and b are the limits of integration.

? 31

Let $f(x)$ be a function ¹⁸For example, $f(x) = \frac{1}{6}x^3 - \frac{1}{2}x^2 + (1+x)\log|x+1|$ will do, but you don't need to know what $f(x)$ is for this problem. with $f''(x) = \frac{x^2}{x+1}$.

1. Show that $|f''(x)| \leq 1$ whenever x is in the interval $[0, 1]$.
2. Find the maximum value of $|f''(x)|$ over the interval $[0, 1]$.
3. Assuming $M = 1$, how many intervals should you use to approximate $\int_0^1 f(x) dx$ to within 10^{-5} ?
4. Using the value of M you found in (b), how many intervals should you use to approximate $\int_0^1 f(x) dx$ to within 10^{-5} ?

? 32

Approximate the function $\log x$ with a rational function by approximating the integral $\int_1^x \frac{1}{t} dt$ using Simpson's rule. Your rational function $f(x)$ should approximate $\log x$ with an error of not more than 0.1 for any x in the interval $[1, 3]$.

? 33

Using an approximation of the area under the curve $\frac{1}{x^2+1}$, show that the constant $\arctan 2$ is in the interval $\left[\frac{\pi}{4} + 0.321, \frac{\pi}{4} + 0.323\right]$.

You may assume use without proof that $\frac{d^4}{dx^4} \left\{ \frac{1}{1+x^2} \right\} = \frac{24(5x^4 - 10x^2 + 1)}{(x^2+1)^5}$. You may use a calculator, but only to add, subtract, multiply, and divide.

1. We apologise for being a little sloppy here — but we just want to say that it can be very hard or even impossible to write some integrals as some finite sized expression involving polynomials, exponentials, logarithms and trigonometric functions. We don't want to get into a discussion of computability, though that is a very interesting topic.
2. Thankfully it is very easy to write a program to apply the midpoint rule.
3. This method is also called the “trapezoid rule” and “trapezium rule”.
4. A trapezoid is a four sided polygon, like a rectangle. But, unlike a rectangle, the top and bottom of a trapezoid need not be parallel.
5. Simpson's rule is named after the 18th century English mathematician Thomas Simpson, despite its use a century earlier by the German mathematician and astronomer Johannes Kepler. In many German texts the rule is often called Kepler's rule.
6. There is an easy generalisation of Simpson's rule that uses cubics instead of parabolas. It is known as Simpson's second rule and Simpson's $\frac{3}{8}$ rule. While one can push this approach further (using quartics, quintics etc), it can sometimes lead to larger errors — the interested reader should look up Runge's phenomenon.
7. Indeed, even beyond the “real world” of many applications in first year calculus texts, some of the methods we have described are used by actual people (such as ship builders, engineers and surveyors) to estimate areas and volumes of actual objects!
8. There is a variant of this trick that works even when you don't know the answer to the integral ahead of time. Suppose that you suspect that the approximation satisfies $M_n = A + K\frac{1}{n^k}$ where A is the exact value of the integral and suppose that you don't know the values of A , K and k . Then $M_n - M_{2n} = K\frac{1}{n^k} - K\frac{1}{(2n)^k} = K\left(1 - \frac{1}{2^k}\right)\frac{1}{n^k}$ so plotting $y = \log(M_n - M_{2n})$ against $x = \log n$ gives the straight line $y = \log \left[K\left(1 - \frac{1}{2^k}\right) \right] - kx$.

9. Now is a good time for a quick revision of logarithms — see “Whirlwind review of logarithms” in Section 2.7 of the CLP-1 text.
10. Linear regression is not part of this course as its derivation requires some multivariable calculus. It is a very standard technique in statistics.
11. This is our favourite running example of an integral that cannot be evaluated algebraically — we need to use numerical methods.
12. The authors tested this empirically.
13. We chose this interval so that we didn't have lots of subscripts floating around in the algebra.
14. The triangle inequality says that for any real numbers x, y $|x + y| \leq |x| + |y|$.

This page titled [1.11: Numerical Integration](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.12: Improper Integrals

Definitions

To this point we have only considered nicely behaved integrals $\int_a^b f(x) dx$. Though the algebra involved in some of our examples was quite difficult, all the integrals had

- finite limits of integration a and b , and
- a bounded integrand $f(x)$ (and in fact continuous except possibly for finitely many jump discontinuities).

Not all integrals we need to study are quite so nice.

Definition 1.12.1

An integral having either an infinite limit of integration or an unbounded integrand is called an improper integral.

Two examples are

$$\int_0^{\infty} \frac{dx}{1+x^2} \quad \text{and} \quad \int_0^1 \frac{dx}{x}$$

The first has an infinite domain of integration and the integrand of the second tends to ∞ as x approaches the left end of the domain of integration. We'll start with an example that illustrates the traps that you can fall into if you treat such integrals sloppily. Then we'll see how to treat them carefully.

✓ Example 1.12.2 $\int_{-1}^1 \frac{1}{x^2} dx$

Consider the integral

$$\int_{-1}^1 \frac{1}{x^2} dx$$

If we “do” this integral completely naively then we get

$$\begin{aligned} \int_{-1}^1 \frac{1}{x^2} dx &= \left. \frac{x^{-1}}{-1} \right|_{-1}^1 \\ &= \frac{1}{-1} - \frac{-1}{-1} \\ &= -2 \end{aligned}$$

which is **wrong**¹. In fact, the answer is ridiculous. The integrand $\frac{1}{x^2} > 0$, so the integral has to be positive.

The flaw in the argument is that the fundamental theorem of calculus, which says that

$$\left| \text{if } F'(x) = f(x) \text{ then } \int_a^b f(x) dx = F(b) - F(a) \right.$$

is applicable only when $F'(x)$ exists and equals $f(x)$ for *all* $a \leq x \leq b$. In this case $F'(x) = \frac{1}{x^2}$ does not exist for $x = 0$. The given integral is improper. We'll see later that the correct answer is $+\infty$.

Let us put this example to one side for a moment and turn to the integral $\int_a^{\infty} \frac{dx}{1+x^2}$. In this case, the integrand is bounded but the domain of integration extends to $+\infty$. We can evaluate this integral by sneaking up on it. We compute it on a bounded domain of integration, like $\int_a^R \frac{dx}{1+x^2}$, and then take the limit $R \rightarrow \infty$.



Let us put this into practice:

✓ Example 1.12.3 $\int_a^\infty \frac{dx}{1+x^2}$

Solution:

- Since the domain extends to $+\infty$ we first integrate on a finite domain

$$\begin{aligned} \int_a^R \frac{dx}{1+x^2} &= \arctan x \Big|_a^R \\ &= \arctan R - \arctan a \end{aligned}$$

- We then take the limit as $R \rightarrow +\infty$:

$$\begin{aligned} \int_a^\infty \frac{dx}{1+x^2} &= \lim_{R \rightarrow \infty} \int_a^R \frac{dx}{1+x^2} \\ &= \lim_{R \rightarrow \infty} [\arctan R - \arctan a] \\ &= \frac{\pi}{2} - \arctan a. \end{aligned}$$

To be more precise, we actually formally *define* an integral with an infinite domain as the limit of the integral with a finite domain as we take one or more of the limits of integration to infinity.

 Definition 1.12.4 Improper integral with infinite domain of integration

1. If the integral $\int_a^R f(x) dx$ exists for all $R > a$, then

$$\int_a^\infty f(x) dx = \lim_{R \rightarrow \infty} \int_a^R f(x) dx$$

when the limit exists (and is finite).

2. If the integral $\int_r^b f(x) dx$ exists for all $r < b$, then

$$\int_{-\infty}^b f(x) dx = \lim_{r \rightarrow -\infty} \int_r^b f(x) dx$$

when the limit exists (and is finite).

3. If the integral $\int_r^R f(x) dx$ exists for all $r < R$, then

$$\int_{-\infty}^\infty f(x) dx = \lim_{r \rightarrow -\infty} \int_r^c f(x) dx + \lim_{R \rightarrow \infty} \int_c^R f(x) dx$$

when both limits exist (and are finite). Any c can be used.

When the limit(s) exist, the integral is said to be convergent. Otherwise it is said to be divergent.

We must also be able to treat an integral like $\int_0^1 \frac{dx}{x}$ that has a finite domain of integration but whose integrand is unbounded near one limit of integration.² Our approach is similar — we sneak up on the problem. We compute the integral on a smaller domain, such as $\int_t^1 \frac{dx}{x}$, with $t > 0$, and then take the limit $t \rightarrow 0^+$.

✓ Example 1.12.5 $\int_0^1 \frac{1}{x} dx$

Solution:

- Since the integrand is unbounded near $x = 0$, we integrate on the smaller domain $t \leq x \leq 1$ with $t > 0$:

$$\int_t^1 \frac{1}{x} dx = \log |x| \Big|_t^1 = -\log |t|$$

- We then take the limit as $t \rightarrow 0^+$ to obtain

$$\int_0^1 \frac{1}{x} dx = \lim_{t \rightarrow 0^+} \int_t^1 \frac{1}{x} dx = \lim_{t \rightarrow 0^+} -\log |t| = +\infty$$

Thus this integral diverges to $+\infty$.

Indeed, we *define* integrals with unbounded integrands via this process:

 **Definition 1.12.6** Improper integral with unbounded integrand

1. If the integral $\int_t^b f(x) dx$ exists for all $a < t < b$, then

$$\int_a^b f(x) dx = \lim_{t \rightarrow a^+} \int_t^b f(x) dx$$

when the limit exists (and is finite).

2. If the integral $\int_a^T f(x) dx$ exists for all $a < T < b$, then

$$\int_a^b f(x) dx = \lim_{T \rightarrow b^-} \int_a^T f(x) dx$$

when the limit exists (and is finite).

3. Let $a < c < b$. If the integrals $\int_a^T f(x) dx$ and $\int_t^b f(x) dx$ exist for all $a < T < c$ and $c < t < b$, then

$$\int_a^b f(x) dx = \lim_{T \rightarrow c^-} \int_a^T f(x) dx + \lim_{t \rightarrow c^+} \int_t^b f(x) dx$$

when both limit exist (and are finite).

When the limit(s) exist, the integral is said to be convergent. Otherwise it is said to be divergent.

Notice that (c) is used when the integrand is unbounded at some point in the middle of the domain of integration, such as was the case in our original example

$$\int_{-1}^1 \frac{1}{x^2} dx$$

A quick computation shows that this integral diverges to $+\infty$

$$\begin{aligned} \int_{-1}^1 \frac{1}{x^2} dx &= \lim_{a \rightarrow 0^-} \int_{-1}^a \frac{1}{x^2} dx + \lim_{b \rightarrow 0^+} \int_b^1 \frac{1}{x^2} dx \\ &= \lim_{a \rightarrow 0^-} \left[1 - \frac{1}{a} \right] + \lim_{b \rightarrow 0^+} \left[\frac{1}{b} - 1 \right] \\ &= +\infty \end{aligned}$$

More generally, if an integral has more than one “source of impropriety” (for example an infinite domain of integration and an integrand with an unbounded integrand or multiple infinite discontinuities) then you split it up into a sum of integrals with a single “source of impropriety” in each. For the integral, as a whole, to converge every term in that sum has to converge.

For example

 **Example 1.12.7** $\int_{-\infty}^{\infty} \frac{dx}{(x-2)x^2}$

Consider the integral

$$\int_{-\infty}^{\infty} \frac{dx}{(x-2)x^2}$$

- The domain of integration that extends to both $+\infty$ and $-\infty$.
- The integrand is singular (i.e. becomes infinite) at $x = 2$ and at $x = 0$.
- So we would write the integral as

$$\int_{-\infty}^{\infty} \frac{dx}{(x-2)x^2} = \int_{-\infty}^a \frac{dx}{(x-2)x^2} + \int_a^0 \frac{dx}{(x-2)x^2} + \int_0^b \frac{dx}{(x-2)x^2} + \int_b^2 \frac{dx}{(x-2)x^2} + \int_2^c \frac{dx}{(x-2)x^2} + \int_c^{\infty} \frac{dx}{(x-2)x^2}$$

where

- a is any number strictly less than 0,
- b is any number strictly between 0 and 2, and
- c is any number strictly bigger than 2.

So, for example, take $a = -1, b = 1, c = 3$.

- When we examine the right-hand side we see that
 - the first integral has domain of integration extending to $-\infty$
 - the second integral has an integrand that becomes unbounded as $x \rightarrow 0^-$,
 - the third integral has an integrand that becomes unbounded as $x \rightarrow 0^+$,
 - the fourth integral has an integrand that becomes unbounded as $x \rightarrow 2^-$,
 - the fifth integral has an integrand that becomes unbounded as $x \rightarrow 2^+$, and
 - the last integral has domain of integration extending to $+\infty$.
- Each of these integrals can then be expressed as a limit of an integral on a small domain.

Examples

With the more formal definitions out of the way, we are now ready for some (important) examples.

✓ Example 1.12.8 $\int_1^{\infty} \frac{dx}{x^p}$ with $p > 0$

Solution:

- Fix any $p > 0$.
- The domain of the integral $\int_1^{\infty} \frac{dx}{x^p}$ extends to $+\infty$ and the integrand $\frac{1}{x^p}$ is continuous and bounded on the whole domain.
- So we write this integral as the limit

$$\int_1^{\infty} \frac{dx}{x^p} = \lim_{R \rightarrow \infty} \int_1^R \frac{dx}{x^p}$$

- The antiderivative of $1/x^p$ changes when $p = 1$, so we will split the problem into three cases, $p > 1$, $p = 1$ and $p < 1$.
- When $p > 1$,

$$\begin{aligned} \int_1^R \frac{dx}{x^p} &= \left. \frac{1}{1-p} x^{1-p} \right|_1^R \\ &= \frac{R^{1-p} - 1}{1-p} \end{aligned}$$

Taking the limit as $R \rightarrow \infty$ gives

$$\begin{aligned} \int_1^{\infty} \frac{dx}{x^p} &= \lim_{R \rightarrow \infty} \int_1^R \frac{dx}{x^p} \\ &= \lim_{R \rightarrow \infty} \frac{R^{1-p} - 1}{1-p} \\ &= \frac{-1}{1-p} = \frac{1}{p-1} \end{aligned}$$

since $1 - p < 0$.

- Similarly when $p < 1$ we have

$$\int_1^{\infty} \frac{dx}{x^p} = \lim_{R \rightarrow \infty} \int_1^R \frac{dx}{x^p} = \lim_{R \rightarrow \infty} \frac{R^{1-p} - 1}{1-p} = +\infty$$

because $1 - p > 0$ and the term R^{1-p} diverges to $+\infty$.

- Finally when $p = 1$

$$\int_1^R \frac{dx}{x} = \log |R| - \log 1 = \log R$$

Then taking the limit as $R \rightarrow \infty$ gives us

$$\int_1^{\infty} \frac{dx}{x^p} = \lim_{R \rightarrow \infty} \log |R| = +\infty.$$

- So summarising, we have

$$\int_1^{\infty} \frac{dx}{x^p} = \begin{cases} \text{divergent} & \text{if } p \leq 1 \\ \frac{1}{p-1} & \text{if } p > 1 \end{cases}$$

✓ Example 1.12.9 $\int_0^1 \frac{dx}{x^p}$ with $p > 0$

Solution:

- Again fix any $p > 0$.
- The domain of integration of the integral $\int_0^1 \frac{dx}{x^p}$ is finite, but the integrand $\frac{1}{x^p}$ becomes unbounded as x approaches the left end, 0, of the domain of integration.
- So we write this integral as

$$\int_0^1 \frac{dx}{x^p} = \lim_{t \rightarrow 0^+} \int_t^1 \frac{dx}{x^p}$$

- Again, the antiderivative changes at $p = 1$, so we split the problem into three cases.
- When $p > 1$ we have

$$\begin{aligned} \int_t^1 \frac{dx}{x^p} &= \frac{1}{1-p} x^{1-p} \Big|_t^1 \\ &= \frac{1 - t^{1-p}}{1-p} \end{aligned}$$

Since $1 - p < 0$ when we take the limit as $t \rightarrow 0$ the term t^{1-p} diverges to $+\infty$ and we obtain

$$\int_0^1 \frac{dx}{x^p} = \lim_{t \rightarrow 0^+} \frac{1 - t^{1-p}}{1-p} = +\infty$$

- When $p = 1$ we similarly obtain

$$\begin{aligned} \int_0^1 \frac{dx}{x} &= \lim_{t \rightarrow 0^+} \int_t^1 \frac{dx}{x} \\ &= \lim_{t \rightarrow 0^+} (-\log |t|) \\ &= +\infty \end{aligned}$$

- Finally, when $p < 1$ we have

$$\begin{aligned} \int_0^1 \frac{dx}{x^p} &= \lim_{t \rightarrow 0^+} \int_t^1 \frac{dx}{x^p} \\ &= \lim_{t \rightarrow 0^+} \frac{1 - t^{1-p}}{1-p} = \frac{1}{1-p} \end{aligned}$$

since $1 - p > 0$.

- In summary

$$\int_0^1 \frac{dx}{x^p} = \begin{cases} \frac{1}{1-p} & \text{if } p < 1 \\ \text{divergent} & \text{if } p \geq 1 \end{cases}$$

✓ Example 1.12.10 $\int_0^\infty \frac{dx}{x^p}$ with $p > 0$

Solution:

- Yet again fix $p > 0$.
- This time the domain of integration of the integral $\int_0^\infty \frac{dx}{x^p}$ extends to $+\infty$, and in addition the integrand $\frac{1}{x^p}$ becomes unbounded as x approaches the left end, 0, of the domain of integration.
- So we split the domain in two — given our last two examples, the obvious place to cut is at $x = 1$:

$$\int_0^\infty \frac{dx}{x^p} = \int_0^1 \frac{dx}{x^p} + \int_1^\infty \frac{dx}{x^p}$$

- We saw, in Example 1.12.9, that the first integral diverged whenever $p \geq 1$, and we also saw, in Example 1.12.8, that the second integral diverged whenever $p \leq 1$.
- So the integral $\int_0^\infty \frac{dx}{x^p}$ diverges for all values of p .

✓ Example 1.12.11 $\int_{-1}^1 \frac{dx}{x}$

This is a pretty subtle example. Look at the sketch below:

This suggests that the signed area to the left of the y -axis should exactly cancel the area to the right of the y -axis making the value of the integral $\int_{-1}^1 \frac{dx}{x}$ exactly zero.

But both of the integrals

$$\begin{aligned} \int_0^1 \frac{dx}{x} &= \lim_{t \rightarrow 0^+} \int_t^1 \frac{dx}{x} = \lim_{t \rightarrow 0^+} [\log x]_t^1 = \lim_{t \rightarrow 0^+} \log \frac{1}{t} = +\infty \\ \int_{-1}^0 \frac{dx}{x} &= \lim_{T \rightarrow 0^-} \int_{-1}^T \frac{dx}{x} = \lim_{T \rightarrow 0^-} [\log |x|]_{-1}^T = \lim_{T \rightarrow 0^-} \log |T| = -\infty \end{aligned}$$

diverge so $\int_{-1}^1 \frac{dx}{x}$ diverges. Don't make the mistake of thinking that $\infty - \infty = 0$. **It is undefined.** And it is undefined for good reason.

For example, we have just seen that the area to the right of the y -axis is

$$\lim_{t \rightarrow 0^+} \int_t^1 \frac{dx}{x} = +\infty$$

and that the area to the left of the y -axis is (substitute $-7t$ for T above)

$$\lim_{t \rightarrow 0^+} \int_{-1}^{-7t} \frac{dx}{x} = -\infty$$

If $\infty - \infty = 0$, the following limit should be 0.

$$\begin{aligned} \lim_{t \rightarrow 0^+} \left[\int_t^1 \frac{dx}{x} + \int_{-1}^{-7t} \frac{dx}{x} \right] &= \lim_{t \rightarrow 0^+} \left[\log \frac{1}{t} + \log |-7t| \right] \\ &= \lim_{t \rightarrow 0^+} \left[\log \frac{1}{t} + \log(7t) \right] \\ &= \lim_{t \rightarrow 0^+} \left[-\log t + \log 7 + \log t \right] = \lim_{t \rightarrow 0^+} \log 7 \\ &= \log 7 \end{aligned}$$

This appears to give $\infty - \infty = \log 7$. Of course the number 7 was picked at random. You can make $\infty - \infty$ be any number at all, by making a suitable replacement for 7.

✓ Example 1.12.12 Example 1.12.2 revisited

The careful computation of the integral of Example 1.12.2 is

$$\begin{aligned} \int_{-1}^1 \frac{1}{x^2} dx &= \lim_{T \rightarrow 0^-} \int_{-1}^T \frac{1}{x^2} dx + \lim_{t \rightarrow 0^+} \int_t^1 \frac{1}{x^2} dx \\ &= \lim_{T \rightarrow 0^-} \left[-\frac{1}{x} \right]_{-1}^T + \lim_{t \rightarrow 0^+} \left[-\frac{1}{x} \right]_t^1 \\ &= \infty + \infty \end{aligned}$$

Hence the integral diverges to $+\infty$.

✓ Example 1.12.13 $\int_{-\infty}^{\infty} \frac{dx}{1+x^2}$

Since

$$\begin{aligned} \lim_{R \rightarrow \infty} \int_0^R \frac{dx}{1+x^2} &= \lim_{R \rightarrow \infty} \left[\arctan x \right]_0^R = \lim_{R \rightarrow \infty} \arctan R = \frac{\pi}{2} \\ \lim_{r \rightarrow -\infty} \int_r^0 \frac{dx}{1+x^2} &= \lim_{r \rightarrow -\infty} \left[\arctan x \right]_r^0 = \lim_{r \rightarrow -\infty} -\arctan r = \frac{\pi}{2} \end{aligned}$$

The integral $\int_{-\infty}^{\infty} \frac{dx}{1+x^2}$ converges and takes the value π .

✓ Example 1.12.14 When does $\int_e^{\infty} \frac{dx}{x(\log x)^p}$ converge?

For what values of p does $\int_e^{\infty} \frac{dx}{x(\log x)^p}$ converge?

Solution:

- For $x \geq e$, the denominator $x(\log x)^p$ is never zero. So the integrand is bounded on the entire domain of integration and this integral is improper only because the domain of integration extends to $+\infty$ and we proceed as usual.
- We have

$$\begin{aligned} \int_e^{\infty} \frac{dx}{x(\log x)^p} &= \lim_{R \rightarrow \infty} \int_e^R \frac{dx}{x(\log x)^p} && \text{use substitution} \\ &= \lim_{R \rightarrow \infty} \int_1^{\log R} \frac{du}{u^p} && \text{with } u = \log x, du = \frac{dx}{x} \\ &= \lim_{R \rightarrow \infty} \begin{cases} \frac{1}{1-p} [(\log R)^{1-p} - 1] & \text{if } p \neq 1 \\ \log(\log R) & \text{if } p = 1 \end{cases} \\ &= \begin{cases} \text{divergent} & \text{if } p \leq 1 \\ \frac{1}{p-1} & \text{if } p > 1 \end{cases} \end{aligned}$$

In this last step we have used similar logic that that used in Example 1.12.8, but with R replaced by $\log R$.

✓ Example 1.12.15 The gamma function

The gamma function $\Gamma(x)$ is defined by the improper integral

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx$$

We shall now compute $\Gamma(n)$ for all natural numbers n .

- To get started, we'll compute

$$\Gamma(1) = \int_0^{\infty} e^{-x} dx = \lim_{R \rightarrow \infty} \int_0^R e^{-x} dx = \lim_{R \rightarrow \infty} \left[-e^{-x} \right]_0^R = 1$$

- Then compute

$$\begin{aligned} \Gamma(2) &= \int_0^{\infty} x e^{-x} dx \\ &= \lim_{R \rightarrow \infty} \int_0^R x e^{-x} dx \end{aligned}$$

Use integration by parts with $u = x$, $dv = e^{-x} dx$, so $v = -e^{-x}$, $du = dx$

$$\begin{aligned} &= \lim_{R \rightarrow \infty} \left[-x e^{-x} \Big|_0^R + \int_0^R e^{-x} dx \right] \\ &= \lim_{R \rightarrow \infty} \left[-x e^{-x} - e^{-x} \right]_0^R \\ &= 1 \end{aligned}$$

For the last equality, we used that $\lim_{x \rightarrow \infty} x e^{-x} = 0$.

- Now we move on to general n , using the same type of computation as we just used to evaluate $\Gamma(2)$. For any natural number n ,

$$\begin{aligned} \Gamma(n+1) &= \int_0^{\infty} x^n e^{-x} dx \\ &= \lim_{R \rightarrow \infty} \int_0^R x^n e^{-x} dx \end{aligned}$$

Again integrate by parts with $u = x^n$, $dv = e^{-x} dx$, so $v = -e^{-x}$, $du = nx^{n-1} dx$

$$\begin{aligned} &= \lim_{R \rightarrow \infty} \left[-x^n e^{-x} \Big|_0^R + \int_0^R nx^{n-1} e^{-x} dx \right] \\ &= \lim_{R \rightarrow \infty} n \int_0^R x^{n-1} e^{-x} dx \\ &= n\Gamma(n) \end{aligned}$$

To get to the third row, we used that $\lim_{x \rightarrow \infty} x^n e^{-x} = 0$.

- Now that we know $\Gamma(2) = 1$ and $\Gamma(n+1) = n\Gamma(n)$, for all $n \in \mathbb{N}$, we can compute all of the $\Gamma(n)$'s.

$$\begin{aligned} \Gamma(2) &= 1 \\ \Gamma(3) &= \Gamma(2+1) = 2\Gamma(2) = 2 \cdot 1 \\ \Gamma(4) &= \Gamma(3+1) = 3\Gamma(3) = 3 \cdot 2 \cdot 1 \\ \Gamma(5) &= \Gamma(4+1) = 4\Gamma(4) = 4 \cdot 3 \cdot 2 \cdot 1 \\ &\vdots \\ \Gamma(n) &= (n-1) \cdot (n-2) \cdots 4 \cdot 3 \cdot 2 \cdot 1 = (n-1)! \end{aligned}$$

That is, the factorial is just ³ the Gamma function shifted by one.

Convergence Tests for Improper Integrals

It is very common to encounter integrals that are too complicated to evaluate explicitly. Numerical approximation schemes, evaluated by computer, are often used instead (see Section 1.11). You want to be sure that at least the integral converges before feeding it into a computer ⁴. Fortunately it is usually possible to determine whether or not an improper integral converges even when you cannot evaluate it explicitly.

Remark 1.12.16

For pedagogical purposes, we are going to concentrate on the problem of determining whether or not an integral $\int_a^\infty f(x) dx$ converges, when $f(x)$ has no singularities for $x \geq a$. Recall that the first step in analyzing any improper integral is to write it as a sum of integrals each of which has only a single “source of impropriety” — either a domain of integration that extends to $+\infty$, or a domain of integration that extends to $-\infty$, or an integrand which is singular at one end of the domain of integration. So we are now going to consider only the first of these three possibilities. But the techniques that we are about to see have obvious analogues for the other two possibilities.

Now let's start. Imagine that we have an improper integral $\int_a^\infty f(x) dx$, that $f(x)$ has no singularities for $x \geq a$ and that $f(x)$ is complicated enough that we cannot evaluate the integral explicitly⁵. The idea is to find another improper integral $\int_a^\infty g(x) dx$

- with $g(x)$ simple enough that we can evaluate the integral $\int_a^\infty g(x) dx$ explicitly, or at least determine easily whether or not $\int_a^\infty g(x) dx$ converges, and
- with $g(x)$ behaving enough like $f(x)$ for large x that the integral $\int_a^\infty f(x) dx$ converges if and only if $\int_a^\infty g(x) dx$ converges.

So far, this is a pretty vague strategy. Here is a theorem which starts to make it more precise.

Theorem 1.12.17 Comparison

Let a be a real number. Let f and g be functions that are defined and continuous for all $x \geq a$ and assume that $g(x) \geq 0$ for all $x \geq a$.

1. If $|f(x)| \leq g(x)$ for all $x \geq a$ and if $\int_a^\infty g(x) dx$ converges then $\int_a^\infty f(x) dx$ also converges.
2. If $f(x) \geq g(x)$ for all $x \geq a$ and if $\int_a^\infty g(x) dx$ diverges then $\int_a^\infty f(x) dx$ also diverges.

We will not prove this theorem, but, hopefully, the following supporting arguments should at least appear reasonable to you. Consider the figure below:

- If $\int_a^\infty g(x) dx$ converges, then the area of

$$\{ (x, y) \mid x \geq a, 0 \leq y \leq g(x) \} \text{ is finite.}$$

When $|f(x)| \leq g(x)$, the region

$$\{ (x, y) \mid x \geq a, 0 \leq y \leq |f(x)| \} \text{ is contained inside } \{ (x, y) \mid x \geq a, 0 \leq y \leq g(x) \}$$

and so must also have finite area. Consequently the areas of both the regions

$$\{ (x, y) \mid x \geq a, 0 \leq y \leq f(x) \} \text{ and } \{ (x, y) \mid x \geq a, f(x) \leq y \leq 0 \}$$

are finite too⁶.

- If $\int_a^\infty g(x) dx$ diverges, then the area of

$$\{ (x, y) \mid x \geq a, 0 \leq y \leq g(x) \} \text{ is infinite.}$$

When $f(x) \geq g(x)$, the region

$$\{ (x, y) \mid x \geq a, 0 \leq y \leq f(x) \} \text{ contains the region } \{ (x, y) \mid x \geq a, 0 \leq y \leq g(x) \}$$

and so also has infinite area.

Example 1.12.18 $\int_1^\infty e^{-x^2} dx$

We cannot evaluate the integral $\int_1^\infty e^{-x^2} dx$ explicitly⁷, however we would still like to understand if it is finite or not — does it converge or diverge?

Solution: We will use Theorem 1.12.17 to answer the question.

- So we want to find another integral that we can compute and that we can compare to $\int_1^\infty e^{-x^2} dx$. To do so we pick an integrand that looks like e^{-x^2} , but whose indefinite integral we know — such as e^{-x} .

- When $x \geq 1$, we have $x^2 \geq x$ and hence $e^{-x^2} \leq e^{-x}$. Thus we can use Theorem 1.12.17 to compare

$$\int_1^{\infty} e^{-x^2} dx \text{ with } \int_1^{\infty} e^{-x} dx$$

- The integral

$$\begin{aligned} \int_1^{\infty} e^{-x} dx &= \lim_{R \rightarrow \infty} \int_1^R e^{-x} dx \\ &= \lim_{R \rightarrow \infty} \left[-e^{-x} \right]_1^R \\ &= \lim_{R \rightarrow \infty} \left[e^{-1} - e^{-R} \right] = e^{-1} \end{aligned}$$

converges.

- So, by Theorem 1.12.17, with $a = 1$, $f(x) = e^{-x^2}$ and $g(x) = e^{-x}$, the integral $\int_1^{\infty} e^{-x^2} dx$ converges too (it is approximately equal to 0.1394).

✓ Example 1.12.19 $\int_{1/2}^{\infty} e^{-x^2} dx$

Solution:

- The integral $\int_{1/2}^{\infty} e^{-x^2} dx$ is quite similar to the integral $\int_1^{\infty} e^{-x^2} dx$ of Example 1.12.18. But we cannot just repeat the argument of Example 1.12.18 because it is not true that $e^{-x^2} \leq e^{-x}$ when $0 < x < 1$.
- In fact, for $0 < x < 1$, $x^2 < x$ so that $e^{-x^2} > e^{-x}$.
- However the difference between the current example and Example 1.12.18 is

$$\int_{1/2}^{\infty} e^{-x^2} dx - \int_1^{\infty} e^{-x^2} dx = \int_{1/2}^1 e^{-x^2} dx$$

which is clearly a well defined finite number (its actually about 0.286). It is important to note that we are being a little sloppy by taking the difference of two integrals like this — we are assuming that both integrals converge. More on this below.

- So we would expect that $\int_{1/2}^{\infty} e^{-x^2} dx$ should be the sum of the proper integral $\int_{1/2}^1 e^{-x^2} dx$ and the convergent integral $\int_1^{\infty} e^{-x^2} dx$ and so should be a convergent integral. This is indeed the case. The Theorem below provides the justification.

🔗 Theorem 1.12.20

Let a and c be real numbers with $a < c$ and let the function $f(x)$ be continuous for all $x \geq a$. Then the improper integral $\int_a^{\infty} f(x) dx$ converges if and only if the improper integral $\int_c^{\infty} f(x) dx$ converges.

Proof

By definition the improper integral $\int_a^{\infty} f(x) dx$ converges if and only if the limit

$$\begin{aligned} \lim_{R \rightarrow \infty} \int_a^R f(x) dx &= \lim_{R \rightarrow \infty} \left[\int_a^c f(x) dx + \int_c^R f(x) dx \right] \\ &= \int_a^c f(x) dx + \lim_{R \rightarrow \infty} \int_c^R f(x) dx \end{aligned}$$

exists and is finite. (Remember that, in computing the limit, $\int_a^c f(x) dx$ is a finite constant independent of R and so can be pulled out of the limit.) But that is the case if and only if the limit $\lim_{R \rightarrow \infty} \int_c^R f(x) dx$ exists and is finite, which in turn is the case if and only if the integral $\int_c^{\infty} f(x) dx$ converges.

✓ Example 1.12.21 Does $\int_1^\infty \frac{\sqrt{x}}{x^2+x} dx$ converge?

Does the integral $\int_1^\infty \frac{\sqrt{x}}{x^2+x} dx$ converge or diverge?

Solution:

- Our first task is to identify the potential sources of impropriety for this integral.
- The domain of integration extends to $+\infty$, but we must also check to see if the integrand contains any singularities. On the domain of integration $x \geq 1$ so the denominator is never zero and the integrand is continuous. So the only problem is at $+\infty$.
- Our second task is to develop some intuition⁸. As the only problem is that the domain of integration extends to infinity, whether or not the integral converges will be determined by the behavior of the integrand for very large x .
- When x is very large, x^2 is much much larger than x (which we can write as $x^2 \gg x$) so that the denominator $x^2 + x \approx x^2$ and the integrand

$$\frac{\sqrt{x}}{x^2 + x} \approx \frac{\sqrt{x}}{x^2} = \frac{1}{x^{3/2}}$$

- By Example 1.12.8, with $p = \frac{3}{2}$, the integral $\int_1^\infty \frac{dx}{x^{3/2}}$ converges. So we would expect that $\int_1^\infty \frac{\sqrt{x}}{x^2+x} dx$ converges too.
- Our final task is to verify that our intuition is correct. To do so, we want to apply part (a) of Theorem 1.12.17 with $f(x) = \frac{\sqrt{x}}{x^2+x}$ and $g(x)$ being $\frac{1}{x^{3/2}}$, or possibly some constant times $\frac{1}{x^{3/2}}$. That is, we need to show that for all $x \geq 1$ (i.e. on the domain of integration)

$$\frac{\sqrt{x}}{x^2 + x} \leq \frac{A}{x^{3/2}}$$

for some constant A . Let's try this.

- Since $x \geq 1$ we know that

$$x^2 + x > x^2$$

Now take the reciprocal of both sides:

$$\frac{1}{x^2 + x} < \frac{1}{x^2}$$

Multiply both sides by \sqrt{x} (which is always positive, so the sign of the inequality does not change)

$$\frac{\sqrt{x}}{x^2 + x} < \frac{\sqrt{x}}{x^2} = \frac{1}{x^{3/2}}$$

- So Theorem 1.12.17(a) and Example 1.12.8, with $p = \frac{3}{2}$ do indeed show that the integral $\int_1^\infty \frac{\sqrt{x}}{x^2+x} dx$ converges.

Notice that in this last example we managed to show that the integral exists by finding an integrand that behaved the same way for large x . Our intuition then had to be bolstered with some careful inequalities to apply the comparison Theorem 1.12.17. It would be nice to avoid this last step and be able jump from the intuition to the conclusion without messing around with inequalities. Thankfully there is a variant of Theorem 1.12.17 that is often easier to apply and that also fits well with the sort of intuition that we developed to solve Example 1.12.21.

A key phrase in the previous paragraph is “behaves the same way for large x ”. A good way to formalise this expression — “ $f(x)$ behaves like $g(x)$ for large x ” — is to require that the limit

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} \text{ exists and is a finite nonzero number.}$$

Suppose that this is the case and call the limit $L \neq 0$. Then

- the ratio $\frac{f(x)}{g(x)}$ must approach L as x tends to $+\infty$.
- So when x is very large — say $x > B$, for some big number B — we must have that

$$\frac{1}{2}L \leq \frac{f(x)}{g(x)} \leq 2L \quad \text{for all } x > B$$

Equivalently, $f(x)$ lies between $\frac{L}{2}g(x)$ and $2Lg(x)$, for all $x \geq B$.

- Consequently, the integral of $f(x)$ converges if and only if the integral of $g(x)$ converges, by Theorems 1.12.17 and 1.12.20.

These considerations lead to the following variant of Theorem 1.12.17.

Theorem 1.12.22 Limiting comparison

Let $-\infty < a < \infty$. Let f and g be functions that are defined and continuous for all $x \geq a$ and assume that $g(x) \geq 0$ for all $x \geq a$.

1. If $\int_a^\infty g(x) dx$ converges and the limit

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)}$$

exists, then $\int_a^\infty f(x) dx$ converges.

2. If $\int_a^\infty g(x) dx$ diverges and the limit

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)}$$

exists and **is nonzero**, then $\int_a^\infty f(x) dx$ diverges.

Note that in (b) the limit must exist and be nonzero, while in (a) we only require that the limit exists (it can be zero).

Here is an example of how Theorem 1.12.22 is used.

Example 1.12.23 $\int_1^\infty \frac{x + \sin x}{e^{-x} + x^2} dx$

Does the integral $\int_1^\infty \frac{x + \sin x}{e^{-x} + x^2} dx$ converge or diverge?

Solution:

- Our first task is to identify the potential sources of impropriety for this integral.
- The domain of integration extends to $+\infty$. On the domain of integration the denominator is never zero so the integrand is continuous. Thus the only problem is at $+\infty$.
- Our second task is to develop some intuition about the behavior of the integrand for very large x . A good way to start is to think about the size of each term when x becomes big.
- When x is very large:
 - $e^{-x} \ll x^2$, so that the denominator $e^{-x} + x^2 \approx x^2$, and
 - $|\sin x| \leq 1 \ll x$, so that the numerator $x + \sin x \approx x$, and
 - the integrand $\frac{x + \sin x}{e^{-x} + x^2} \approx \frac{x}{x^2} = \frac{1}{x}$.

Notice that we are using $A \ll B$ to mean that “ A is much much smaller than B ”. Similarly $A \gg B$ means “ A is much much bigger than B ”. We don't really need to be too precise about its meaning beyond this in the present context.

- Now, since $\int_1^\infty \frac{dx}{x}$ diverges, we would expect $\int_1^\infty \frac{x + \sin x}{e^{-x} + x^2} dx$ to diverge too.
- Our final task is to verify that our intuition is correct. To do so, we set

$$f(x) = \frac{x + \sin x}{e^{-x} + x^2} \quad g(x) = \frac{1}{x}$$

and compute

$$\begin{aligned}\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} &= \lim_{x \rightarrow \infty} \frac{x + \sin x}{e^{-x} + x^2} \div \frac{1}{x} \\ &= \lim_{x \rightarrow \infty} \frac{(1 + \sin x/x)x}{(e^{-x}/x^2 + 1)x^2} \times x \\ &= \lim_{x \rightarrow \infty} \frac{1 + \sin x/x}{e^{-x}/x^2 + 1} \\ &= 1\end{aligned}$$

- Since $\int_1^\infty g(x) dx = \int_1^\infty \frac{dx}{x}$ diverges, by Example 1.12.8 with $p = 1$, Theorem 1.12.22(b) now tells us that $\int_1^\infty f(x) dx = \int_1^\infty \frac{x + \sin x}{e^{-x} + x^2} dx$ diverges too.

Exercises

Stage 1

? 1

For which values of b is the integral $\int_0^b \frac{1}{x^2 - 1} dx$ improper?

? 2

For which values of b is the integral $\int_0^b \frac{1}{x^2 + 1} dx$ improper?

? 3

Below are the graphs $y = f(x)$ and $y = g(x)$. Suppose $\int_0^\infty f(x) dx$ converges, and $\int_0^\infty g(x) dx$ diverges. Assuming the graphs continue on as shown as $x \rightarrow \infty$, which graph is $f(x)$, and which is $g(x)$?

? 4 (*)

Decide whether the following statement is true or false. If false, provide a counterexample. If true, provide a brief justification. (Assume that $f(x)$ and $g(x)$ are continuous functions.)

If $\int_1^\infty f(x) dx$ converges and $g(x) \geq f(x) \geq 0$ for all x , then $\int_1^\infty g(x) dx$ converges.

? 5

Let $f(x) = e^{-x}$ and $g(x) = \frac{1}{x+1}$. Note $\int_0^\infty f(x) dx$ converges while $\int_0^\infty g(x) dx$ diverges.

For each of the functions $h(x)$ described below, decide whether $\int_0^\infty h(x) dx$ converges or diverges, or whether there isn't enough information to decide. Justify your decision.

1. $h(x)$, continuous and defined for all $x \geq 0$, $h(x) \leq f(x)$.
2. $h(x)$, continuous and defined for all $x \geq 0$, $f(x) \leq h(x) \leq g(x)$.
3. $h(x)$, continuous and defined for all $x \geq 0$, $-2f(x) \leq h(x) \leq f(x)$.

Stage 2

? 6 (*)

Evaluate the integral $\int_0^1 \frac{x^4}{x^5 - 1} dx$ or state that it diverges.

? 7 (*)

Determine whether the integral $\int_{-2}^2 \frac{1}{(x+1)^{4/3}} dx$ is convergent or divergent. If it is convergent, find its value.

? 8 (*)

Does the improper integral $\int_1^{\infty} \frac{1}{\sqrt{4x^2 - x}} dx$ converge? Justify your answer.

? 9 (*)

Does the integral $\int_0^{\infty} \frac{dx}{x^2 + \sqrt{x}}$ converge or diverge? Justify your claim.

? 10

Does the integral $\int_{-\infty}^{\infty} \cos x dx$ converge or diverge? If it converges, evaluate it.

? 11

Does the integral $\int_{-\infty}^{\infty} \sin x dx$ converge or diverge? If it converges, evaluate it.

? 12

Evaluate $\int_{10}^{\infty} \frac{x^4 - 5x^3 + 2x - 7}{x^5 + 3x + 8} dx$, or state that it diverges.

? 13

Evaluate $\int_0^{10} \frac{x-1}{x^2 - 11x + 10} dx$, or state that it diverges.

? 14 (*)

Determine (with justification!) which of the following applies to the integral $\int_{-\infty}^{+\infty} \frac{x}{x^2 + 1} dx$:

1. $\int_{-\infty}^{+\infty} \frac{x}{x^2 + 1} dx$ diverges
2. $\int_{-\infty}^{+\infty} \frac{x}{x^2 + 1} dx$ converges but $\int_{-\infty}^{+\infty} \left| \frac{x}{x^2 + 1} \right| dx$ diverges
3. $\int_{-\infty}^{+\infty} \frac{x}{x^2 + 1} dx$ converges, as does $\int_{-\infty}^{+\infty} \left| \frac{x}{x^2 + 1} \right| dx$

Remark: these options, respectively, are that the integral diverges, converges conditionally, and converges absolutely. You'll see this terminology used for series in Section 3.4.1.

? 15 (*)

Decide whether $I = \int_0^{\infty} \frac{|\sin x|}{x^{3/2} + x^{1/2}} dx$ converges or diverges. Justify.

? 16 (*)

Does the integral $\int_0^{\infty} \frac{x+1}{x^{1/3}(x^2+x+1)} dx$ converge or diverge?

Stage 3

? 17

We craft a tall, vuvuzela-shaped solid by rotating the line $y = \frac{1}{x}$ from $x = a$ to $x = 1$ about the y -axis, where a is some constant between 0 and 1.

True or false: **No matter how large a constant M is, there is some value of a that makes a solid with volume larger than M .**

? 18 (*)

What is the largest value of q for which the integral $\int_1^{\infty} \frac{1}{x^{5q}} dx$ diverges?

? 19

For which values of p does the integral $\int_0^{\infty} \frac{x}{(x^2+1)^p} dx$ converge?

? 20

Evaluate $\int_2^{\infty} \frac{1}{t^4-1} dt$, or state that it diverges.

? 21

Does the integral $\int_{-5}^5 \left(\frac{1}{\sqrt{|x|}} + \frac{1}{\sqrt{|x-1|}} + \frac{1}{\sqrt{|x-2|}} \right) dx$ converge or diverge?

? 22

Evaluate $\int_0^{\infty} e^{-x} \sin x dx$, or state that it diverges.

? 23 (*)

Is the integral $\int_0^{\infty} \frac{\sin^4 x}{x^2} dx$ convergent or divergent? Explain why.

? 24

Does the integral $\int_0^{\infty} \frac{x}{e^x + \sqrt{x}} dx$ converge or diverge?

? 25 (*)

Let $M_{n,t}$ be the Midpoint Rule approximation for $\int_0^t \frac{e^{-x}}{1+x} dx$ with n equal subintervals. Find a value of t and a value of n such that $M_{n,t}$ differs from $\int_0^{\infty} \frac{e^{-x}}{1+x} dx$ by at most 10^{-4} . Recall that the error E_n introduced when the Midpoint Rule is used with n subintervals obeys

$$|E_n| \leq \frac{M(b-a)^3}{24n^2}$$

where M is the maximum absolute value of the second derivative of the integrand and a and b are the end points of the interval of integration.

? 26

Suppose $f(x)$ is continuous for all real numbers, and $\int_1^{\infty} f(x) dx$ converges.

1. If $f(x)$ is odd, does $\int_{-\infty}^{-1} f(x) dx$ converge or diverge, or is there not enough information to decide?
2. If $f(x)$ is even, does $\int_{-\infty}^{\infty} f(x) dx$ converge or diverge, or is there not enough information to decide?

? 27

True or false:

There is some real number x , with $x \geq 1$, such that $\int_0^x \frac{1}{e^t} dt = 1$.

1. Very wrong. But it is not an example of “not even wrong” — which is a phrase attributed to the physicist Wolfgang Pauli who was known for his harsh critiques of sloppy arguments. The phrase is typically used to describe arguments that are so incoherent that not only can one not prove they are true, but they lack enough coherence to be able to show they are false. The interested reader should do a little searchengining and look at the concept of falisfyability.
2. This will, in turn, allow us to deal with integrals whose integrand is unbounded somewhere inside the domain of integration.
3. The Gamma function is far more important than just a generalisation of the factorial. It appears all over mathematics, physics, statistics and beyond. It has all sorts of interesting properties and its definition can be extended from natural numbers n to all numbers excluding $0, -1, -2, -3, \dots$. For example, one can show that $\Gamma(1-z)\Gamma(z) = \frac{\pi}{\sin \pi z}$.
4. Applying numerical integration methods to a divergent integral may result in perfectly reasonably looking but very wrong answers.
5. You could, for example, think of something like our running example $\int_a^{\infty} e^{-t^2} dt$.
6. We have separated the regions in which $f(x)$ is positive and negative, because the integral $\int_a^{\infty} f(x) dx$ represents the signed area of the union of $\{(x, y) \mid x \geq a, 0 \leq y \leq f(x)\}$ and $\{(x, y) \mid x \geq a, f(x) \leq y \leq 0\}$.
7. It has been the subject of many remarks and footnotes.
8. This takes practice, practice and more practice. At the risk of alliteration — please perform plenty of practice problems.

This page titled [1.12: Improper Integrals](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.13: More Integration Examples

Exercises

Recall that we are using $\log x$ to denote the logarithm of x with base e . In other courses it is often denoted $\ln x$.

Stage 1

? 1

Match the integration method to a common kind of integrand it's used to antidifferentiate.

(A) $u = f(x)$ substitution	(I)	a function multiplied by its derivative
(B) trigonometric substitution	(II)	a polynomial times an exponential
(C) integration by parts	(III)	a rational function
(D) partial fractions	(IV)	the square root of a quadratic function

Stage 2

? 2

Evaluate $\int_0^{\frac{\pi}{2}} \sin^4 x \cos^5 x \, dx$.

? 3

Evaluate $\int \sqrt{3 - 5x^2} \, dx$.

? 4

Evaluate $\int_0^{\infty} \frac{x-1}{e^x} \, dx$.

? 5

Evaluate $\int \frac{-2}{3x^2 + 4x + 1} \, dx$.

? 6

Evaluate $\int_1^2 x^2 \log x \, dx$.

? 7 (*)

Evaluate $\int \frac{x}{x^2 - 3} \, dx$.

? 8 (*)

Evaluate the following integrals.

1. $\int_0^4 \frac{x}{\sqrt{9+x^2}} dx$
2. $\int_0^{\pi/2} \cos^3 x \sin^2 x dx$
3. $\int_1^e x^3 \log x dx$

? 9 (*)

Evaluate the following integrals.

1. $\int_0^{\pi/2} x \sin x dx$
2. $\int_0^{\pi/2} \cos^5 x dx$

? 10 (*)

Evaluate the following integrals.

1. $\int_0^2 x e^x dx$
2. $\int_0^1 \frac{1}{\sqrt{1+x^2}} dx$
3. $\int_3^5 \frac{4x}{(x^2-1)(x^2+1)} dx$

? 11 (*)

Calculate the following integrals.

1. $\int_0^3 \sqrt{9-x^2} dx$
2. $\int_0^1 \log(1+x^2) dx$
3. $\int_3^\infty \frac{x}{(x-1)^2(x-2)} dx$

? 12

Evaluate $\int \frac{\sin^4 \theta - 5 \sin^3 \theta + 4 \sin^2 \theta + 10 \sin \theta}{\sin^2 \theta - 5 \sin \theta + 6} \cos \theta d\theta$.

? 13 (*)

Evaluate the following integrals. Show your work.

1. $\int_0^{\frac{\pi}{4}} \sin^2(2x) \cos^3(2x) dx$
2. $\int (9+x^2)^{-\frac{3}{2}} dx$
3. $\int \frac{dx}{(x-1)(x^2+1)}$
4. $\int x \arctan x dx$

? 14 (*)

Evaluate the following integrals.

1. $\int_0^{\pi/4} \sin^5(2x) \cos(2x) dx$
2. $\int \sqrt{4-x^2} dx$
3. $\int \frac{x+1}{x^2(x-1)} dx$

? 15 (*)

Calculate the following integrals.

1. $\int_0^{\infty} e^{-x} \sin(2x) dx$
2. $\int_0^{\sqrt{2}} \frac{1}{(2+x^2)^{3/2}} dx$
3. $\int_0^1 x \log(1+x^2) dx$
4. $\int_3^{\infty} \frac{1}{(x-1)^2(x-2)} dx$

? 16 (*)

Evaluate the following integrals.

1. $\int x \log x dx$
2. $\int \frac{(x-1) dx}{x^2+4x+5}$
3. $\int \frac{dx}{x^2-4x+3}$
4. $\int \frac{x^2 dx}{1+x^6}$

? 17 (*)

Evaluate the following integrals.

1. $\int_0^1 \arctan x dx.$
2. $\int \frac{2x-1}{x^2-2x+5} dx. (*)$

? 18 (*)

1. Evaluate $\int \frac{x^2}{(x^3+1)^{101}} dx.$
2. Evaluate $\int \cos^3 x \sin^4 x dx.$

? 19 (*)

Evaluate $\int_{\pi/2}^{\pi} \frac{\cos x}{\sqrt{\sin x}} dx$.

? 20 (*)

Evaluate the following integrals.

1. $\int \frac{e^x}{(e^x + 1)(e^x - 3)} dx$
2. $\int_2^4 \frac{x^2 - 4x + 4}{\sqrt{12 + 4x - x^2}} dx$

? 21 (*)

Evaluate these integrals.

1. $\int \frac{\sin^3 x}{\cos^3 x} dx$
2. $\int_{-2}^2 \frac{x^4}{x^{10} + 16} dx$

? 22

Evaluate $\int x\sqrt{x-1} dx$.

? 23

Evaluate $\int \frac{\sqrt{x^2 - 2}}{x^2} dx$.

You may use that $\int \sec x dx = \log |\sec x + \tan x| + C$.

? 24

Evaluate $\int_0^{\pi/4} \sec^4 x \tan^5 x dx$.

? 25

Evaluate $\int \frac{3x^2 + 4x + 6}{(x+1)^3} dx$.

? 26

Evaluate $\int \frac{1}{x^2 + x + 1} dx$.

? 27

Evaluate $\int \sin x \cos x \tan x dx$.

? 28

Evaluate $\int \frac{1}{x^3+1} dx$.

? 29

Evaluate $\int (3x)^2 \arcsin x dx$.**Stage 3**

? 30

Evaluate $\int_0^{\pi/2} \sqrt{\cos t + 1} dt$.

? 31

Evaluate $\int_0^e \frac{\log \sqrt{x}}{x} dx$.

? 32

Evaluate $\int_{0.1}^{0.2} \frac{\tan x}{\log(\cos x)} dx$.

? 33 (*)

Evaluate these integrals.

1. $\int \sin(\log x) dx$
2. $\int_0^1 \frac{1}{x^2 - 5x + 6} dx$

? 34 (*)

Evaluate (with justification).

1. $\int_0^3 (x+1)\sqrt{9-x^2} dx$
2. $\int \frac{4x+8}{(x-2)(x^2+4)} dx$
3. $\int_{-\infty}^{+\infty} \frac{1}{e^x + e^{-x}} dx$

? 35

Evaluate $\int \sqrt{\frac{x}{1-x}} dx$.

? 36

Evaluate $\int_0^1 e^{2x} e^{e^x} dx$.

? 37

Evaluate $\int \frac{x e^x}{(x+1)^2} dx$.

? 38

Evaluate $\int \frac{x \sin x}{\cos^2 x} dx$.

You may use that $\int \sec x dx = \log |\sec x + \tan x| + C$.

? 39

Evaluate $\int x(x+a)^n dx$, where a and n are constants.

? 40

Evaluate $\int \arctan(x^2) dx$.

This page titled [1.13: More Integration Examples](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

2: Applications of Integration

In the previous chapter we defined the definite integral, based on its interpretation as the area of a region in the xy -plane. We also developed a bunch of theory to help us work with integrals. This abstract definition, and the associated theory, turns out to be extremely useful simply because "areas of regions in the xy -plane" appear in a huge number of different settings, many of which seem superficially not to involve "areas of regions in the xy -plane". Here are some examples.

- The work involved in moving a particle or in pumping a fluid out of a reservoir. See section 2.1.
- The average value of a function. See section 2.2.
- The center of mass of an object. See section 2.3.
- The time dependence of temperature. See section 2.4.
- Radiocarbon dating. See section 2.4.

Let us start with the first of these examples.

[2.1: Work](#)

[2.2: Averages](#)

[2.3: Center of Mass and Torque](#)

[2.4: Separable Differential Equations](#)

This page titled [2: Applications of Integration](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

2.1: Work

While computing areas and volumes are nice mathematical applications of integration we can also use integration to compute quantities of importance in physics and statistics. One such quantity is work. Work is a way of quantifying the amount of energy that is required to act against a force¹. In SI² metric units the force F has units newtons (which are kilogram-metres per second squared), x has units metres and the work W has units joules (which are newton-metres or kilogram-metres squared per second squared).

Definition 2.1.1

The work done by a force $F(x)$ in moving an object from $x = a$ to $x = b$ is

$$W = \int_a^b F(x) dx$$

In particular, if the force is a constant, F , independent of x , the work is $F \cdot (b - a)$.

Here is some motivation for this definition. Consider a particle of mass m moving along the x -axis. Let the position of the particle at time t be $x(t)$. The particle starts at position a at time α , moves to the right, finishing at position $b > a$ at time β . While the particle moves, it is subject to a position-dependent force $F(x)$. Then Newton's law of motion³ says⁴ that force is mass times acceleration

$$m \frac{d^2 x}{dt^2}(t) = F(x(t))$$

Now consider our definition of work above. It tells us that the work done in moving the particle from $x = a$ to $x = b$ is

$$W = \int_a^b F(x) dx$$

However, we know the position as a function of time, so we can substitute $x = x(t)$, $dx = \frac{dx}{dt} dt$ (using Theorem 1.4.6) and rewrite the above integral:

$$W = \int_a^b F(x) dx = \int_{t=\alpha}^{t=\beta} F(x(t)) \frac{dx}{dt} dt$$

Using Newton's second law we can rewrite our integrand:

$$\begin{aligned} &= m \int_{\alpha}^{\beta} \frac{d^2 x}{dt^2} \frac{dx}{dt} dt \\ &= m \int_{\alpha}^{\beta} \frac{dv}{dt} v(t) dt && \text{since } v(t) = \frac{dx}{dt} \\ &= m \int_{\alpha}^{\beta} \frac{d}{dt} \left(\frac{1}{2} v(t)^2 \right) dt \end{aligned}$$

What happened here? By the chain rule, for any function $f(t)$:

$$\frac{d}{dt} \left(\frac{1}{2} f(t)^2 \right) = f(t) f'(t).$$

In the above computation we have used this fact with $f(t) = v(t)$. Now using the fundamental theorem of calculus (Theorem 1.3.1 part 2), we have

$$\begin{aligned} W &= m \int_{\alpha}^{\beta} \frac{d}{dt} \left(\frac{1}{2} v(t)^2 \right) dt \\ &= \frac{1}{2} m v(\beta)^2 - \frac{1}{2} m v(\alpha)^2. \end{aligned}$$

By definition, the function $\frac{1}{2}mv(t)^2$ is the kinetic energy⁵ of the particle at time t . So the work W of Definition 2.1.1 is the change in kinetic energy from the time the particle was at $x = a$ to the time it was at $x = b$.

✓ Example: 2.1.2 Hooke's Law

Imagine that a spring lies along the x -axis. The left hand end is fixed to a wall, but the right hand end lies freely at $x = 0$. So the spring is at its “natural length”.

- Now suppose that we wish to stretch out the spring so that its right hand end is at $x = L$.
- Hooke's Law⁶ says that when a (linear) spring is stretched (or compressed) by x units beyond its natural length, it exerts a force of magnitude kx , where the constant k is the spring constant of that spring.
- In our case, once we have stretched the spring by x units to the right, the spring will be trying to pull back the right hand end by applying a force of magnitude kx directed to the left.
- For us to continue stretching the spring we will have to apply a compensating force of magnitude kx directed to the right. That is, we have to apply the force $F(x) = +kx$.
- So to stretch a spring by L units from its natural length we have to supply the work

$$W = \int_0^L kx \, dx = \frac{1}{2}kL^2$$

✓ Example: 2.1.3 Spring

A spring has a natural length of 0.1m. If a 12N force is needed to keep it stretched to a length of 0.12m, how much work is required to stretch it from 0.12m to 0.15m?

Solution:

In order to answer this question we will need to determine the spring constant and then integrate the appropriate function.

- Our first task is to determine the spring constant k . We are told that when the spring is stretched to a length of 0.12m, i.e. to a length of $0.12 - 0.1 = 0.02$ m beyond its natural length, then the spring generates a force of magnitude 12N.
- Hooke's law states that the force exerted by the spring, when it is stretched by x units, has magnitude kx , so

$$12 = k \cdot 0.02 = k \cdot \frac{2}{100} \quad \text{thus}$$

$$k = 600.$$

- So to stretch the spring
 - from a length of 0.12m, i.e. a length of $x = 0.12 - 0.1 = 0.02$ m beyond its natural length,
 - to a length of 0.15m, i.e. a length of $x = 0.15 - 0.1 = 0.05$ m beyond its natural length,

takes work

$$W = \int_{0.02}^{0.05} kx \, dx = \left[\frac{1}{2}kx^2 \right]_{0.02}^{0.05}$$

$$= 300(0.05^2 - 0.02^2)$$

$$= 0.63\text{J}$$

✓ Example: 2.1.4 Pumping Out a Reservoir

A cylindrical reservoir⁷ of height h and radius r is filled with a fluid of density ρ . We would like to know how much work is required to pump all of the fluid out the top of the reservoir.

Solution: We are going to tackle this problem by applying the standard integral calculus “slice into small pieces” strategy. This is how we computed areas and volumes — slice the problem into small pieces, work out how much each piece contributes, and then add up the contributions using an integral.

- Start by slicing the reservoir (or rather the fluid inside it) into thin, horizontal, cylindrical pancakes, as in the figure above. We proceed by determining how much work is required to pump out this pancake volume of fluid⁸.
- Each pancake is a squat cylinder with thickness dx and circular cross section of radius r and area πr^2 . Hence it has volume $\pi r^2 dx$ and mass $\rho \times \pi r^2 dx$.
- Near the surface of the Earth gravity exerts a downward force of mg on a body of mass m . The constant $g = 9.8 \text{ m/sec}^2$ is called the *standard acceleration due to gravity*⁹. For us to raise the pancake we have to apply a compensating upward force of mg , which, for our pancake, is

$$F = g\rho \times \pi r^2 dx$$

- To remove the pancake at height x from the reservoir we need to raise it to height h . So we have to lift it a distance $h - x$ using the force $F = \pi \rho g r^2 dx$, which takes work $\pi \rho g r^2 (h - x) dx$.
- The total work to empty the whole reservoir is

$$\begin{aligned} W &= \int_0^h \pi \rho g r^2 (h - x) dx = \pi \rho g r^2 \int_0^h (h - x) dx \\ &= \pi \rho g r^2 \left[hx - \frac{x^2}{2} \right]_0^h \\ &= \frac{\pi}{2} \rho g r^2 h^2 \end{aligned}$$

- If we measure lengths in metres and mass in kilograms, then this quantity has units of Joules. If we instead used feet and pounds¹⁰ then this would have units of “foot-pounds”. One foot-pound is equal to 1.355817... Joules.

✓ Example: 2.1.5 Escape Velocity

Suppose that you shoot a probe straight up from the surface of the Earth — at what initial speed must the probe move in order to escape Earth's gravity?

Solution: We determine this by computing how much work must be done in order to escape Earth's gravity. If we assume that all of this work comes from the probe's initial kinetic energy, then we can establish the minimum initial velocity required.

- The work done by gravity when a mass moves from the surface of the Earth to a height h above the surface is

$$W = \int_0^h F(x) dx$$

where $F(x)$ is the gravitational force acting on the mass at height x above the Earth's surface.

- The gravitational force¹¹ of the Earth acting on a particle of mass m at a height x above the surface of the Earth is

$$F = -\frac{GMm}{(R+x)^2},$$

where G is the gravitational constant, M is the mass of the Earth and R is the radius of the Earth. Note that $R + x$ is the distance from the object to the center of the Earth. Additionally, note that this force is negative because gravity acts downward.

- So the work done by gravity on the probe, as it travels from the surface of the Earth to a height h , is

$$\begin{aligned} W &= -\int_0^h \frac{GMm}{(R+x)^2} dx \\ &= -GMm \int_0^h \frac{1}{(R+x)^2} dx \end{aligned}$$

A quick application of the substitution rule with $u = R + x$ gives

$$\begin{aligned}
 &= -GMm \int_{u(0)}^{u(h)} \frac{1}{u^2} du \\
 &= -GMm \left[-\frac{1}{u} \right]_{u=R}^{u=R+h} \\
 &= \frac{GMm}{R+h} - \frac{GMm}{R}
 \end{aligned}$$

- So if the probe completely escapes the Earth and travels all the way to $h = \infty$, gravity does work

$$\lim_{h \rightarrow \infty} \left[\frac{GMm}{R+h} - \frac{GMm}{R} \right] = -\frac{GMm}{R}$$

The minus sign means that gravity has removed energy $\frac{GMm}{R}$ from the probe.

- To finish the problem we need one more assumption. Let us assume that all of this energy comes from the probe's initial kinetic energy and that the probe is not fitted with any sort of rocket engine. Hence the initial kinetic energy $\frac{1}{2}mv^2$ (coming from an initial velocity v) must be at least as large as the work computed above. That is we need

$$\begin{aligned}
 \frac{1}{2}mv^2 &\geq \frac{GMm}{R} && \text{which rearranges to give} \\
 v &\geq \sqrt{\frac{2GM}{R}}
 \end{aligned}$$

- The right hand side of this inequality, $\sqrt{\frac{2GM}{R}}$, is called the escape velocity.

✓ Example: 2.1.6 Lifting a Cable

A 10-metre-long cable of mass 5kg is used to lift a bucket of water, with mass 8kg, out of a well. Find the work done.

Solution: Denote by y the height of the bucket above the top of the water in the well. So the bucket is raised from $y = 0$ to $y = 10$. The cable has mass density 0.5kg/m. So when the bucket is at height y ,

- the cable that remains to be lifted has mass $0.5(10 - y)$ kg and
- the remaining cable and water is subject to a downward gravitational force of magnitude $[0.5(10 - y) + 8]g = [13 - \frac{y}{2}]g$, where $g = 9.8$ m/sec².

So to raise the bucket from height y to height $y + dy$ we need to apply a compensating upward force of $[13 - \frac{y}{2}]g$ through distance dy . This takes work $[13 - \frac{y}{2}]g dy$. So the total work required is

$$\int_0^{10} \left[13 - \frac{y}{2} \right] g dy = g \left[13y - \frac{y^2}{4} \right]_0^{10} = [130 - 25]g = 105g \text{ J}$$

Exercises

Stage 1

? 1

Find the work (in joules) required to lift a 3-gram block of matter a height of 10 centimetres against the force of gravity (with $g = 9.8$ m/sec²).

? 2

A rock exerts a force of 1 N on the ground where it sits due to gravity. Use $g = 9.8$ m/sec².

What is the mass of the rock?

How much work (in joules) does it take to lift that rock one metre in the air?

? 3

Consider the equation

$$W = \int_a^b F(x) dx$$

where x is measured in metres and $F(x)$ is measured in kilogram-metres per second squared (newtons).

For some large n , we might approximate

$$W \approx \sum_{i=1}^n F(x_i) \Delta x$$

where $\Delta x = \frac{b-a}{n}$ and x_i is some number in the interval $[a + (i-1)\Delta x, a + i\Delta x]$. (This is just the general form of a Riemann sum).

1. What are the units of Δx ?
2. What are the units of $F(x_i)$?
3. Using your answers above, what are the units of W ?

Remark: we already know the units of W from the text, but the Riemann sum illustrates *why they make sense* arising from this particular integral.

? 4

Suppose $f(x)$ has units $\frac{\text{smoot}}{\text{megaFonzie}}$, and x is measured in barns¹². What are the units of the quantity $\int_0^1 f(x) dx$?

? 5

You want to weigh your luggage before a flight. You don't have a scale or balance, but you do have a heavy-duty spring from your local engineering-supply store. You nail it to your wall, marking where the bottom hangs. You hang a one-litre bag of water (with mass one kilogram) from the spring, and observe that the spring stretches 1 cm. Where on the wall should you mark the bottom of the spring corresponding to a hanging mass of 10kg?

You may assume that the spring obeys Hooke's law.

? 6

The work done by a force in moving an object from position $x = 1$ to $x = b$ is $W(b) = -b^3 + 6b^2 - 9b + 4$ for any b in $[1, 3]$. At what position x in $[1, 3]$ is the force the strongest?

Stage 2

Questions 9 through 16 offer practice on two broad types of calculations covered in the text: lifting things against gravity, and stretching springs. You may make the same physical assumptions as in the text: that is, springs follow Hooke's law, and the acceleration due to gravity is a constant -9.8 metres per second squared.

For Questions 18 and 19, use the principle (introduced after Definition 2.1.1 and utilized in Example 2.1.5) that the work done on a particle by a force over a distance is equal to the change in kinetic energy of that particle.

? 7 (*)

A variable force $F(x) = \frac{a}{\sqrt{x}}$ Newtons moves an object along a straight line when it is a distance of x meters from the origin. If the work done in moving the object from $x = 1$ meters to $x = 16$ meters is 18 joules, what is the value of a ? Don't worry about the units of a .

? 8

A tube of air is fitted with a plunger that compresses the air as it is pushed in. If the natural length of the tube of air is ℓ , when the plunger has been pushed x metres past its natural position, the force exerted by the air is $\frac{c}{\ell-x}$ N, where c is a positive constant (depending on the particulars of the tube of air) and $x < \ell$.

1. What are the units of c ?
2. How much work does it take to push the plunger from 1 metre past its natural position to 1.5 metres past its natural position? (You may assume $\ell > 1.5$.)

? 9 (*)

Find the work (in joules) required to stretch a string 10 cm beyond equilibrium, if its spring constant is $k = 50$ N/m.

? 10 (*)

A force of 10 N (newtons) is required to hold a spring stretched 5 cm beyond its natural length. How much work, in joules (J), is done in stretching the spring from its natural length to 50 cm beyond its natural length?

? 11 (*)

A 5-metre-long cable of mass 8 kg is used to lift a bucket off the ground. How much work is needed to raise the entire cable to height 5 m? Ignore the mass of the bucket and its contents.

? 12

A tank 1 metre high has pentagonal cross sections of area 3 m^2 and is filled with water. How much work does it take to pump out all the water?

You may assume the density of water is 1 kg per 1000 cm^3 .

? 13 (*)

A sculpture, shaped like a pyramid 3m high sitting on the ground, has been made by stacking smaller and smaller (very thin) iron plates on top of one another. The iron plate at height z m above ground level is a square whose side length is $(3 - z)$ m. All of the iron plates started on the floor of a basement 2 m below ground level.

Write down an integral that represents the work, in joules, it took to move all of the iron from its starting position to its present position. Do *not* evaluate the integral. (You can use 9.8 m/s^2 for the acceleration due to gravity and 8000 kg/m^3 for the density of iron.)

? 14

Suppose a spring extends 5 cm past its natural length when one kilogram is hung from its end. How much work is done to extend the spring from 5 cm past its natural length to 7 cm past its natural length?

? 15

Ten kilograms of firewood are hoisted on a rope up a height of 4 metres to a second-floor deck. If the total work done is 400 joules, what is the mass of the 4 metres of rope?

You may assume that the rope has the same density all the way along.

? 16

A 5 kg weight is attached to the middle of a 10-metre long rope, which dangles out a window. The rope alone has mass 1 kg. How much work does it take to pull the entire rope in through the window, together with the weight?

? 17

A box is dragged along the floor. Friction exerts a force in the opposite direction of motion from the box, and that force is equal to $\mu \times m \times g$, where μ is a constant, m is the mass of the box and g is the acceleration due to gravity. You may assume $g = 9.8 \text{ m/sec}^2$.

1. How much work is done dragging a box of mass 10 kg along the floor for three metres if $\mu = 0.4$?
2. Suppose the box contains a volatile substance that rapidly evaporates. You pull the box at a constant rate of 1 m/sec for three seconds, and the mass of the box at t seconds ($0 \leq t \leq 3$) is $(10 - \sqrt{t})$ kilograms. If $\mu = 0.4$, how much work is done pulling the box for three seconds?

? 18

A ball of mass 1 kg is attached to a spring, and the spring is attached to a table. The ball moves with some initial velocity, and the spring slows it down. At its farthest, the spring stretches 10 cm past its natural length. If the spring constant is 5 N/m, what was the initial velocity of the ball?

You may assume that the ball starts moving with initial velocity v_0 , and that the only force slowing it down is the spring. You may also assume that the spring started out at its natural length, it follows Hooke's law, and when it is stretched its farthest, the velocity of the ball is 0 m/sec.

? 19

A mild-mannered university professor who is definitely not a spy notices that when their car is on the ground, it is 2 cm shorter than when it is on a jack. (That is: when the car is on a jack, its struts are at their natural length; when on the ground, the weight of the car causes the struts to compress 2 cm.) The university professor calculates that if they were to jump a local neighborhood drawbridge, their car would fall to the ground with a speed of 4 m/sec. If the car can sag 20 cm before important parts scrape the ground, and the car has mass 2000 kg unoccupied (2100 kg with the professor inside), can the professor, who is certainly not involved in international intrigue, safely jump the bridge?

Assume the car falls vertically, the struts obey Hooke's law, and the work done by the struts is equal to the change in kinetic energy of the car + professor. Use 9.8 m/sec^2 for the acceleration due to gravity.

Stage 3

? 20

A disposable paper cup has the shape of a right circular cone with radius 5 cm and height 15 cm, and is completely filled with water. How much work is done sucking all the water out of the cone with a straw?

You may assume that 1 m^3 of water has mass 1000 kilograms, the acceleration due to gravity is -9.8 m/sec^2 , and that the water moves as high up as the very top of the cup and no higher.

? 21 (*)

A spherical tank of radius 3 metres is half-full of water. It has a spout of length 1 metre sticking up from the top of the tank. Find the work required to pump all of the water in the tank out the spout. The density of water is 1000 kilograms per cubic metre. The acceleration due to gravity is $9.8 \text{ metres per second squared}$.

? 22

A 5-metre cable is pulled out of a deep hole, where it was dangling straight down. The cable has density $\rho(x) = (10 - x)$ kg/m, where x is the distance from the bottom end of the rope. (So, the bottom of the cable is denser than the top.) How much work is done pulling the cable out of the hole?

? 23

A rectangular tank is fitted with a plunger that can raise and lower the water level by decreasing and increasing the length of its base, as in the diagrams below. The tank has base width 1 m (which does not change) and contains 3 m^3 of water.

The force of the water acting on any tiny piece of the plunger is PA , where P is the pressure of the water, and dA is the area of the tiny piece. The pressure varies with the depth of the piece (below the surface of the water). Specifically, $P = cD$, where D is the depth of the tiny piece and c is a constant, in this case $c = 9800 \text{ N/m}^3$.

1. If the length of the base is 3 m, give the force of the water on the entire plunger. (You can do this with an integral: it's the sum of the force on all the tiny pieces of the plunger.)
2. If the length of the base is x m, give the force of the water on the entire plunger.
3. Give the work required to move the plunger in so that the base length changes from 3 m to 1 m.

? 24

A leaky bucket picks up 5 L of water from a well, but drips out 1 L every ten seconds. If the bucket was hauled up 5 metres at a constant speed of 1 metre every two seconds, how much work was done?

Assume the rope and bucket have negligible mass and one litre of water has 1 kg mass, and use 9.8 m/sec^2 for the acceleration due to gravity.

? 25

The force of gravity between two objects, one of mass m_1 and another of mass m_2 , is $F = G \frac{m_1 m_2}{r^2}$, where r is the distance between them and G is the gravitational constant.

How much work is required to separate the earth and the moon far enough apart that the gravitational attraction between them is negligible?

Assume the mass of the earth is 6×10^{24} kg and the mass of the moon is 7×10^{22} kg, and that they are currently 400 000 km away from each other. Also, assume $G = 6.7 \times 10^{-11} \frac{\text{m}^3}{\text{kg} \cdot \text{sec}^2}$, and the only force acting on the earth and moon is the gravity between them.

? 26

True or false: the work done pulling up a dangling cable of length ℓ and mass m (with uniform density) is the same as the work done lifting up a ball of mass m a height of $\ell/2$.

? 27

A tank one metre high is filled with watery mud that has settled to be denser at the bottom than at the top.

At height h metres above the bottom of the tank, the cross-section of the tank has the shape of the finite region bounded by the two curves $y = x^2$ and $y = 2 - h - 3x^2$. At height h metres above the bottom of the tank, the density of the liquid is $1000\sqrt{2 - h}$ kilograms per cubic metre.

How much work is done to pump all the liquid out of the tank?

You may assume the acceleration due to gravity is 9.8 m/sec^2 .

? 28

An hourglass is 0.2 m tall and shaped such that that y metres above or below its vertical center it has a radius of $y^2 + 0.01$ m. It is exactly half-full of sand, which has mass $M = \frac{1}{7}$ kilograms.

How much work is done on the sand by quickly flipping the hourglass over?

Assume that the work done is only moving against gravity, with $g = 9.8$ m/sec², and the sand has uniform density. Also assume that at the instant the hourglass is flipped over, the sand has not yet begun to fall, as in the picture above.

? 29

Suppose at position x a particle experiences a force of $F(x) = \sqrt{1 - x^4}$ N. Approximate the work done moving the particle from $x = 0$ to $x = 1/2$, accurate to within 0.01 J.

1. For example — if your expensive closed-source textbook has fallen on the floor, work quantifies the amount of energy required to lift the object from the floor acting against the force of gravity.
2. SI is short for “le système international d’unités” which is French for “the international system of units”. It is the most recent internationally sanctioned version of the metric system, published in 1960. It aims to establish sensible units of measurement (no cubic furlongs per hogshead-Fahrenheit). It defines seven base units — metre (length), kilogram (mass), second (time), kelvin (temperature), ampere (electric current), mole (quantity of substance) and candela (luminous intensity). From these one can then establish derived units — such as metres per second for velocity and speed.
3. Specifically, the second of Newton's three law of motion. These were first published in 1687 in his “Philosophiæ Naturalis Principia Mathematica”.
4. It actually says something more graceful in Latin - Mutationem motus proportionalem esse vi motrici impressae, et fieri secundum lineam rectam qua vis illa imprimitur. Or — The alteration of motion is ever proportional to the motive force impressed; and is made in the line in which that force is impressed. It is amazing what you can find on the internet.
5. This is not a physics text so we will not be too precise. Roughly speaking, kinetic energy is the energy an object possesses due to it being in motion, as opposed to potential energy, which is the energy of the object due to its position in a force field. Leibniz and Bernoulli determined that kinetic energy is proportional to the square of the velocity, while the modern term “kinetic energy” was first used by Lord Kelvin (back while he was still William Thompson).
6. Robert Hooke (1635–1703) was an English contemporary of Isaac Newton (1643–1727). It was in a 1676 letter to Hooke that Newton wrote “If I have seen further it is by standing on the shoulders of Giants.” There is some thought that this was sarcasm and Newton was actually making fun of Hooke, who had a spinal deformity. However at that time Hooke and Newton were still friends. Several years later they did have a somewhat public falling-out over some of Newton's work on optics.
7. We could assign units to these measurements — such as metres for the lengths h and r , and kilograms per cubic metre for the density ρ .
8. Potential for a bad “work out how much work out” pun here.
9. This quantity is not actually constant — it varies slightly across the surface of earth depending on local density, height above sea-level and centrifugal force from the earth's rotation. It is, for example, slightly higher in Oslo and slightly lower in Singapore. It is actually *defined* to be 9.80665m/sec² by the International Organisation for Standardization.
10. It is extremely mysterious to both authors why a person would do science or engineering in imperial units. One of the authors still has nightmares about having had to do so as a student. The other author is thankful that he escaped such tortures.
11. Newton published his inverse square law of universal gravitation in his Principia in 1687. His law states that the gravitational force between two masses m_1 and m_2 is $F = -G \frac{m_1 m_2}{r^2}$ where r is the distance separating the (centers of the) masses and $G = 6.674 \times 10^{-11} \text{Nm}^2/\text{kg}^2$ is the gravitational constant. Notice that r measures the separation between the centers of the masses not the distance between the surfaces of the objects. Also, do not confuse G with g — standard acceleration due to gravity. The first measurement of G was performed by Henry Cavendish in 1798 — the interested reader should look up the “Cavendish experiment” for details of this very impressive work.
12. For this problem, it doesn't matter what the units measure, but a smoot is a silly measure of length; a megaFonzie is an apocryphal measure of coolness; and a barn is a humorous (but actually used) measure of area. For explanations (and

entertainment) see https://en.Wikipedia.org/wiki/List_of_humorous_units_of_measurement and https://en.Wikipedia.org/wiki/List_of_unusual_units_of_measurement (accessed 27 July 2017).

This page titled [2.1: Work](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

2.2: Averages

Another frequent ¹ application of integration is computing averages and other statistical quantities. We will not spend too much time on this topic — that is best left to a proper course in statistics — however, we will demonstrate the application of integration to the problem of computing averages.

Let us start with the definition ²

$$\begin{aligned} \text{arithmetic mean} &= \frac{1}{n}(y_1 + y_2 + \cdots + y_n) \\ \text{geometric mean} &= (y_1 \cdot y_2 \cdots y_n)^{\frac{1}{n}} \\ \text{harmonic mean} &= \left[\frac{1}{n} \left(\frac{1}{y_1} + \frac{1}{y_2} + \cdots + \frac{1}{y_n} \right) \right]^{-1} \end{aligned}$$

All of these quantities, along with the median and mode, are ways to measure the typical value of a set of numbers. They all have advantages and disadvantages — another interesting topic beyond the scope of this course, but plenty of fodder for the interested reader and their favourite search engine. But let us put pedantry (and beyond-the-scope-of-the-course-reading) aside and just use the terms average and mean interchangeably for our purposes here. of the average of a finite set of numbers.

Definition: 2.2.1

The average (mean) of a set of n numbers y_1, y_2, \dots, y_n is

$$y_{\text{ave}} = \bar{y} = \langle y \rangle = \frac{y_1 + y_2 + \cdots + y_n}{n}$$

The notations y_{ave} , \bar{y} and $\langle y \rangle$ are all commonly used to represent the average.

Now suppose that we want to take the average of a function $f(x)$ with x running continuously from a to b . How do we even define what that means? A natural approach is to

- select, for each natural number n , a sample of n , more or less uniformly distributed, values of x between a and b ,
- take the average of the values of f at the selected points,
- and then take the limit as n tends to infinity.

Unsurprisingly, this process looks very much like how we computed areas and volumes previously. So let's get to it.

- First fix any natural number n .
- Subdivide the interval $a \leq x \leq b$ into n equal subintervals, each of width $\Delta x = \frac{b-a}{n}$.
- The subinterval number i runs from x_{i-1} to x_i with $x_i = a + i \frac{b-a}{n}$.
- Select, for each $1 \leq i \leq n$, one value of x from subinterval number i and call it x_i^* . So $x_{i-1} \leq x_i^* \leq x_i$.
- The average value of f at the selected points is

$$\frac{1}{n} \sum_{i=1}^n f(x_i^*) = \frac{1}{b-a} \sum_{i=1}^n f(x_i^*) \Delta x \quad \text{since } \Delta x = \frac{b-a}{n}$$

giving us a Riemann sum.

Now when we take the limit $n \rightarrow \infty$ we get exactly $\frac{1}{b-a} \int_a^b f(x) dx$. That's why we define

Definition: 2.2.2

Let $f(x)$ be an integrable function defined on the interval $a \leq x \leq b$. The average value of f on that interval is

$$f_{\text{ave}} = \bar{f} = \langle f \rangle = \frac{1}{b-a} \int_a^b f(x) dx$$

Consider the case when $f(x)$ is positive. Then rewriting Definition 2.2.2 as

$$f_{\text{ave}}(b-a) = \int_a^b f(x) dx$$

gives us a link between the average value and the area under the curve. The right-hand side is the area of the region

$$\{(x, y) \mid a \leq x \leq b, 0 \leq y \leq f(x)\}$$

while the left-hand side can be seen as the area of a rectangle of width $b-a$ and height f_{ave} . Since these areas must be the same, we interpret f_{ave} as the height of the rectangle which has the same width and the same area as $\{(x, y) \mid a \leq x \leq b, 0 \leq y \leq f(x)\}$.

Let us start with a couple of simple examples and then work our way up to harder ones.

✓ Example: 2.2.3 An easy warm-up

Let $f(x) = x$ and $g(x) = x^2$ and compute their average values over $1 \leq x \leq 5$.

Solution: We can just plug things into the definition.

$$\begin{aligned} f_{\text{ave}} &= \frac{1}{5-1} \int_1^5 x dx \\ &= \frac{1}{4} \left[\frac{x^2}{2} \right]_1^5 \\ &= \frac{1}{8} (25 - 1) = \frac{24}{8} \\ &= 3 \end{aligned}$$

as we might expect. And then

$$\begin{aligned} g_{\text{ave}} &= \frac{1}{5-1} \int_1^5 x^2 dx \\ &= \frac{1}{4} \left[\frac{x^3}{3} \right]_1^5 \\ &= \frac{1}{12} (125 - 1) = \frac{124}{12} \\ &= \frac{31}{3} \end{aligned}$$

Something a little more trigonometric

✓ Example: 2.2.4 Average of sine

Find the average value of $\sin(x)$ over $0 \leq x \leq \frac{\pi}{2}$.

Solution: Again, we just need the definition.

$$\begin{aligned} \text{average} &= \frac{1}{\frac{\pi}{2} - 0} \int_0^{\frac{\pi}{2}} \sin(x) dx \\ &= \frac{2}{\pi} \cdot \left[-\cos(x) \right]_0^{\frac{\pi}{2}} \\ &= \frac{2}{\pi} (-\cos(\frac{\pi}{2}) + \cos(0)) \\ &= \frac{2}{\pi}. \end{aligned}$$

We could keep going... But better to do some more substantial examples.

✓ Example: 2.2.5 Average velocity

Let $x(t)$ be the position at time t of a car moving along the x -axis. The velocity of the car at time t is the derivative $v(t) = x'(t)$. The average velocity of the car over the time interval $a \leq t \leq b$ is

$$\begin{aligned} v_{\text{ave}} &= \frac{1}{b-a} \int_a^b v(t) dt \\ &= \frac{1}{b-a} \int_a^b x'(t) dt \\ &= \frac{x(b) - x(a)}{b-a} \end{aligned} \quad \text{by the fundamental theorem of calculus.}$$

The numerator in this formula is just the displacement (net distance travelled — if $x'(t) \geq 0$, it's the distance travelled) between time a and time b and the denominator is just the time it took.

Notice that this is exactly the formula we used way back at the start of your *differential* calculus class to help introduce the idea of the derivative. Of course this is a very circuitous way to get to this formula — but it is reassuring that we get the same answer.

A very physics example.

✓ Example: 2.2.6 Peak vs RMS voltage

When you plug a light bulb into a socket³ and turn it on, it is subjected to a voltage

$$V(t) = V_0 \sin(\omega t - \delta)$$

where

- $V_0 = 170$ volts,
- $\omega = 2\pi \times 60$ (which corresponds to 60 cycles per second⁴) and
- the constant δ is an (unimportant) phase. It just shifts the time at which the voltage is zero

The voltage V_0 is the “peak voltage” — the maximum value the voltage takes over time. More typically we quote the “root mean square” voltage⁵ (or RMS-voltage). In this example we explain the difference, but to simplify the calculations, let us simplify the voltage function and just use

$$V(t) = V_0 \sin(t)$$

Since the voltage is a sine-function, it takes both positive and negative values. If we take its simple average over 1 period then we get

$$\begin{aligned} V_{\text{ave}} &= \frac{1}{2\pi - 0} \int_0^{2\pi} V_0 \sin(t) dt \\ &= \frac{V_0}{2\pi} \left[-\cos(t) \right]_0^{2\pi} \\ &= \frac{V_0}{2\pi} (-\cos(2\pi) + \cos(0)) = \frac{V_0}{2\pi} (-1 + 1) \\ &= 0 \end{aligned}$$

This is clearly not a good indication of the typical voltage.

What we actually want here is a measure of how far the voltage is from zero. Now we could do this by taking the average of $|V(t)|$, but this is a little harder to work with. Instead we take the average of the square⁶

$$\text{quadratic mean} = \sqrt{\frac{1}{n} (y_1^2 + y_2^2 + \cdots + y_n^2)}$$

of the voltage (so it is always positive) and then take the square root at the end. That is

$$\begin{aligned}
 V_{\text{rms}} &= \sqrt{\frac{1}{2\pi - 0} \int_0^{2\pi} V(t)^2 dt} \\
 &= \sqrt{\frac{1}{2\pi} \int_0^{2\pi} V_0^2 \sin^2(t) dt} \\
 &= \sqrt{\frac{V_0^2}{2\pi} \int_0^{2\pi} \sin^2(t) dt}
 \end{aligned}$$

This is called the “root mean square” voltage.

Though we do know how to integrate sine and cosine, we don't (yet) know how to integrate their squares. A quick look at double-angle formulas⁷ gives us a way to eliminate the square:

$$\cos(2\theta) = 1 - 2 \sin^2 \theta \implies \sin^2 \theta = \frac{1 - \cos(2\theta)}{2}$$

Using this we manipulate our integrand a little more:

$$\begin{aligned}
 V_{\text{rms}} &= \sqrt{\frac{V_0^2}{2\pi} \int_0^{2\pi} \frac{1}{2} (1 - \cos(2t)) dt} \\
 &= \sqrt{\frac{V_0^2}{4\pi} \left[t - \frac{1}{2} \sin(2t) \right]_0^{2\pi}} \\
 &= \sqrt{\frac{V_0^2}{4\pi} \left(2\pi - \frac{1}{2} \sin(4\pi) - 0 + \frac{1}{2} \sin(0) \right)} \\
 &= \sqrt{\frac{V_0^2}{4\pi} \cdot 2\pi} \\
 &= \frac{V_0}{\sqrt{2}}
 \end{aligned}$$

So if the peak voltage is 170 volts then the RMS voltage is $\frac{170}{\sqrt{2}} \approx 120.2$.

Continuing this very physics example:

✓ Example: 2.2.7 Peak vs RMS voltage — continued

Let us take our same light bulb with voltage (after it is plugged in) given by

$$V(t) = V_0 \sin(\omega t - \delta)$$

where

- V_0 is the peak voltage,
- $\omega = 2\pi \times 60$, and
- the constant δ is an (unimportant) phase.

If the light bulb is “100 watts”, then what is its resistance?

To answer this question we need the following facts from physics.

- If the light bulb has resistance R ohms, this causes, by Ohm's law, a current of

$$I(t) = \frac{1}{R} V(t)$$

(amps) to flow through the light bulb.

- The current I is the number of units of charge moving through the bulb per unit time.
- The voltage is the energy required to move one unit of charge through the bulb.

- The power is the energy used by the bulb per unit time and is measured in watts.

So the power is the product of the current times the voltage and, so

$$P(t) = I(t)V(t) = \frac{V(t)^2}{R} = \frac{V_0^2}{R} \sin^2(\omega t - \delta)$$

The average power used over the time interval $a \leq t \leq b$ is

$$P_{\text{ave}} = \frac{1}{b-a} \int_a^b P(t) dt = \frac{V_0^2}{R(b-a)} \int_a^b \sin^2(\omega t - \delta) dt$$

Notice that this is almost exactly the form we had in the previous example when computing the root mean square voltage.

Again we simplify the integrand using the identity

$$\cos(2\theta) = 1 - 2 \sin^2 \theta \implies \sin^2 \theta = \frac{1 - \cos(2\theta)}{2}$$

So

$$\begin{aligned} P_{\text{ave}} &= \frac{1}{b-a} \int_a^b P(t) dt = \frac{V_0^2}{2R(b-a)} \int_a^b [1 - \cos(2\omega t - 2\delta)] dt \\ &= \frac{V_0^2}{2R(b-a)} \left[t - \frac{\sin(2\omega t - 2\delta)}{2\omega} \right]_a^b \\ &= \frac{V_0^2}{2R(b-a)} \left[b - a - \frac{\sin(2\omega b - 2\delta)}{2\omega} + \frac{\sin(2\omega a - 2\delta)}{2\omega} \right] \\ &= \frac{V_0^2}{2R} - \frac{V_0^2}{4\omega R(b-a)} [\sin(2\omega b - 2\delta) - \sin(2\omega a - 2\delta)] \end{aligned}$$

In the limit as the length of the time interval $b - a$ tends to infinity, this converges to $\frac{V_0^2}{2R}$. The resistance R of a “100 watt bulb” obeys

$$\frac{V_0^2}{2R} = 100 \quad \text{so that} \quad R = \frac{V_0^2}{200}.$$

We finish this example off with two side remarks.

- If we translate the peak voltage to the root mean square voltage using

$$V_0 = V_{\text{rms}} \cdot \sqrt{2}$$

then we have

$$P = \frac{V_{\text{rms}}^2}{R}$$

- If we were using direct voltage rather than alternating current then the computation is much simpler. The voltage and current are constants, so

$$\begin{aligned} P &= V \cdot I \quad \text{but } I = V/R \text{ by Ohm's law} \\ &= \frac{V^2}{R} \end{aligned}$$

So if we have a direct current giving voltage equal to the root mean square voltage, then we would expend the same power.

Optional — Return to the mean value theorem

Here is another application of the Definition 2.2.2 of the average value of a function on an interval. The following theorem can be thought of as an analogue of the mean value theorem (which was covered in your differential calculus class) but for integrals. The theorem says that a continuous function $f(x)$ must be exactly equal to its average value for some x . For example, if you went for a

drive along the x -axis and you were at $x(a)$ at time a and at $x(b)$ at time b , then your velocity $x'(t)$ had to be exactly your average velocity $\frac{x(b)-x(a)}{b-a}$ at some time t between a and b . In particular, if your average velocity was greater than the speed limit, you were definitely speeding at some point during the trip. This is, of course, no great surprise⁸.

Theorem: 2.2.8 Mean Value Theorem for Integrals

Let $f(x)$ be a continuous function on the interval $a \leq x \leq b$. Then there is some c obeying $a < c < b$ such that

$$\frac{1}{b-a} \int_a^b f(x) dx = f(c) \quad \text{or} \quad \int_a^b f(x) dx = f(c)(b-a)$$

Proof

We will apply the mean value theorem (Theorem 2.13.4 in the CLP-1 text) to the function

$$F(x) = \int_a^x f(t) dt$$

By the part 1 of the fundamental theorem of calculus (Theorem 1.3.1), $F'(x) = f(x)$, so the mean value theorem says that there is a $a < c < b$ with

$$\begin{aligned} f(c) = F'(c) &= \frac{F(b) - F(a)}{b-a} = \frac{1}{b-a} \left\{ \int_a^b f(t) dt - \int_a^a f(t) dt \right\} \\ &= \frac{1}{b-a} \int_a^b f(x) dx \end{aligned}$$

In the next section, we will encounter an application in which we want to take the average value of a function $f(x)$, but in doing so we want some values of x to count more than other values of x . That is, we want to weight some x 's more than other x 's. To do so, we choose a "weight function" $w(x) \geq 0$ with $w(x)$ larger for more important x 's. Then we define the weighted average of f as follows.

Definition: 2.2.9

Let $f(x)$ and $w(x)$ be integrable functions defined on the interval $a \leq x \leq b$ with $w(x) \geq 0$ for all $a \leq x \leq b$ and with $\int_a^b w(x) dx > 0$. The average value of f on that interval, weighted by w , is

$$\frac{\int_a^b f(x) w(x) dx}{\int_a^b w(x) dx}$$

We typically refer to this simply as the weighted average of f .

Here are a few remarks concerning this definition.

- The definition has been rigged so that, if $f(x) = 1$ for all x , then the weighted average of f is 1, no matter what weight function $w(x)$ is used.
- If the weight function $w(x) = C$ for some constant $C > 0$ then the weighted average

$$\frac{\int_a^b f(x) w(x) dx}{\int_a^b w(x) dx} = \frac{\int_a^b f(x) C dx}{\int_a^b C dx} = \frac{\int_a^b f(x) dx}{b-a}$$

is just the usual average.

- For any function $w(x) \geq 0$ and any $a < b$, we have $\int_a^b w(x) dx \geq 0$. But for the definition of weighted average to make sense, we need to be able to divide by $\int_a^b w(x) dx$. So we need $\int_a^b w(x) dx \neq 0$.

The next theorem says that a continuous function $f(x)$ must be equal to its weighted average at some point x .

Theorem: 2.2.10 Mean Value Theorem for Weighted Integrals

Let $f(x)$ and $w(x)$ be continuous functions on the interval $a \leq x \leq b$. Assume that $w(x) > 0$ for all $a < x < b$. Then there is some c obeying $a < c < b$ such that

$$\frac{\int_a^b f(x) w(x) dx}{\int_a^b w(x) dx} = f(c) \quad \text{or} \quad \int_a^b f(x) w(x) dx = f(c) \int_a^b w(x) dx$$

Proof

We will apply the generalised mean value theorem (Theorem 3.4.38 in the CLP-1 text) to

$$F(x) = \int_a^x f(t) w(t) dt \quad G(x) = \int_a^x w(t) dt$$

By the part 1 of the fundamental theorem of calculus (Theorem 1.3.1), $F'(x) = f(x)w(x)$ and $G'(x) = w(x)$, so the generalised mean value theorem says that there is a $a < c < b$ with

$$\begin{aligned} f(c) &= \frac{F'(c)}{G'(c)} = \frac{F(b) - F(a)}{G(b) - G(a)} = \frac{\int_a^b f(t) w(t) dt - \int_a^a f(t) w(t) dt}{\int_a^b w(t) dt - \int_a^a w(t) dt} \\ &= \frac{\int_a^b f(t) w(t) dt}{\int_a^b w(t) dt} \end{aligned}$$

✓ Example: 2.2.11

In this example, we will take a number of weighted averages of the simple function $f(x) = x$ over the simple interval $a = 1 \leq x \leq 2 = b$. As x increases from 1 to 2, the function $f(x)$ increases linearly from 1 to 2. So it is no shock that the ordinary average of f is exactly its middle value:

$$\frac{1}{b-a} \int_a^b f(t) dt = \frac{1}{2-1} \int_1^2 t dt = \frac{3}{2}$$

Pick any natural number $N \geq 1$ and consider the weight function $w_N(x) = x^N$. Note that $w_N(x)$ increases as x increases. So $w_N(x)$ weights bigger x 's more than it weights smaller x 's. In particular w_N weights the point $x = 2$ by a factor of 2^N (which is greater than 1 and grows to infinity as N grows to infinity) more than it weights the point $x = 1$. The weighted average of f is

$$\begin{aligned} \frac{\int_a^b f(t) w_N(t) dt}{\int_a^b w_N(t) dt} &= \frac{\int_1^2 t^{N+1} dt}{\int_1^2 t^N dt} = \frac{\frac{2^{N+2}-1}{N+2}}{\frac{2^{N+1}-1}{N+1}} = \frac{N+1}{N+2} \frac{2^{N+2}-1}{2^{N+1}-1} \\ &= \begin{cases} \frac{2 \times 7}{3 \times 3} = 1.555 & \text{if } N = 1 \\ \frac{3 \times 15}{4 \times 7} = 1.607 & \text{if } N = 2 \\ \frac{4 \times 31}{5 \times 15} = 1.653 & \text{if } N = 3 \\ \frac{5 \times 63}{6 \times 31} = 1.694 & \text{if } N = 4 \\ 1.889 & \text{if } N = 16 \\ 1.992 & \text{if } N = 256 \end{cases} \end{aligned}$$

As we would expect, the w_N -weighted average is between 1.5 (which is the ordinary, unweighted, average) and 2 (which is the biggest value of f in the interval) and grows as N grows. The limit as $N \rightarrow \infty$ of the w_N -weighted average is

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{N+1}{N+2} \frac{2^{N+2}-1}{2^{N+1}-1} &= \lim_{N \rightarrow \infty} \frac{N+2-1}{N+2} \frac{2^{N+2}-2+1}{2^{N+1}-1} \\ &= \lim_{N \rightarrow \infty} \left[1 - \frac{1}{N+2} \right] \left[2 + \frac{1}{2^{N+1}-1} \right] \\ &= 2 \end{aligned}$$

✓ Example: 2.2.12

Here is an example which shows what can go wrong with Theorem 2.2.10 if we allow the weight function $w(x)$ to change sign. Let $a = -0.99$ and $b = 1$. Let

$$w(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Then

$$\int_a^b f(x) w(x) dx = \int_0^1 x dx = \frac{1}{2}$$

$$\int_a^b w(x) dx = \int_0^1 dx - \int_{-0.99}^0 dx = 1 - 0.99 = 0.01$$

As c runs from a to b , $f(c) \int_a^b w(x) dx = 0.01 f(c)$ runs from 0 to 0.01 and, in particular, never takes a value anywhere near $\int_a^b f(x) w(x) dx = \frac{1}{2}$. There is no c value which works.

Exercises

Recall that we are using $\log x$ to denote the logarithm of x with base e . In other courses it is often denoted $\ln x$.

Stage 1

? 1

Below is the graph of a function $y = f(x)$. Its average value on the interval $[0, 5]$ is A . Draw a rectangle on the graph with area $\int_0^5 f(x) dx$.

? 2

Suppose a car travels for 5 hours in a straight line, with an average velocity of 100 kph. How far did the car travel?

? 3

A force $F(x)$ acts on an object from position $x = a$ metres to position $x = b$ metres, for a total of W joules of work. What was the average force on the object?

? 4

Suppose we want to approximate the average value of the function $f(x)$ on the interval $[a, b]$. To do this, we cut the interval $[a, b]$ into n pieces, then take n samples by finding the function's output at the left endpoint of each piece, starting with a . Then, we average those n samples. (In the example below, $n = 4$.)

1. Using n samples, what is the distance between two consecutive sample points x_i and x_{i+1} ?
2. Assuming $n \geq 4$, what is the x -coordinate of the fourth sample?
3. Assuming $n \geq 4$, what is the y -value of the fourth sample?
4. Write the approximation of the average value of $f(x)$ over the interval $[a, b]$ using sigma notation.

? 5

Suppose $f(x)$ and $g(x)$ are functions that are defined for all numbers in the interval $[0, 10]$.

1. If $f(x) \leq g(x)$ for all x in $[0, 10]$, then is the average value of $f(x)$ is less than or equal to the average value of $g(x)$ on the interval $[0, 10]$, or is there not enough information to tell?
2. Suppose $f(x) \leq g(x)$ for all x in $[0.01, 10]$. Is the average value of $f(x)$ less than or equal to the average value of $g(x)$ over the interval $[0, 10]$, or is there not enough information to tell?

? 6

Suppose f is an odd function, defined for all real numbers. What is the average of f on the interval $[-10, 10]$?

Stage 2

For Questions 16 through 18, let the root mean square of $f(x)$ on $[a, b]$ be $\sqrt{\frac{1}{b-a} \int_a^b f^2(x) dx}$. This is the formula used in Example 2.2.6 in the text.

? 7 (*)

Find the average value of $f(x) = \sin(5x) + 1$ over the interval $-\pi/2 \leq x \leq \pi/2$.

? 8 (*)

Find the average value of the function $y = x^2 \log x$ on the interval $1 \leq x \leq e$.

? 9 (*)

Find the average value of the function $f(x) = 3 \cos^3 x + 2 \cos^2 x$ on the interval $0 \leq x \leq \frac{\pi}{2}$.

? 10 (*)

Let k be a positive constant. Find the average value of the function $f(x) = \sin(kx)$ on the interval $0 \leq x \leq \pi/k$.

? 11 (*)

The temperature in Celsius in a 3 m long rod at a point x metres from the left end of the rod is given by the function $T(x) = \frac{80}{16-x^2}$. Determine the average temperature in the rod.

? 12 (*)

What is the average value of the function $f(x) = \frac{\log x}{x}$ on the interval $[1, e]$?

? 13 (*)

Find the average value of $f(x) = \cos^2(x)$ over $0 \leq x \leq 2\pi$.

? 14

The carbon dioxide concentration in the air at a particular location over one year is approximated by $C(t) = 400 + 50 \cos\left(\frac{t}{12}\pi\right) + 200 \cos\left(\frac{t}{4380}\pi\right)$ parts per million, where t is measured in hours.

1. What is the average carbon dioxide concentration for that location for that year?
2. What is the average over the first day?

3. Suppose measurements were only made at noon every day: that is, when $t = 12 + 24n$, where n is any whole number between 0 and 364. Then the daily variation would cease: $50 \cos\left(\frac{(12+24n)}{12}\pi\right) = 50 \cos(\pi + 2\pi n) = 50 \cos \pi = -50$. So, the approximation for the concentration of carbon dioxide in the atmosphere might be given as

$$N(t) = 350 + 200 \cos\left(\frac{t}{4380}\pi\right) \text{ ppm}$$

What is the relative error in the yearly average concentration of carbon dioxide involved in using $N(t)$, instead of $C(t)$?

You may assume a day has exactly 24 hours, and a year has exactly 8760 hours.

? 15

Let S be the solid formed by rotating the parabola $y = x^2$ from $x = 0$ to $x = 2$ about the x -axis.

1. What is the average area of the circular cross-sections of S ? Call this value A .
2. What is the volume of S ?
3. What is the volume of a cylinder with circular cross-sectional area A and length 2?

? 16

Let $f(x) = x$.

1. Calculate the average of $f(x)$ over $[-3, 3]$.
2. Calculate the root mean square of $f(x)$ over $[-3, 3]$.

? 17

Calculate the root mean square of $f(x) = \tan x$ over $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$.

? 18

A force acts on a spring, and the spring stretches and contracts. The distance beyond its natural length at time t is $f(t) = \sin(t\pi)$ cm, where t is measured in seconds. The spring constant is 3 N/cm.

1. What is the force exerted by the spring at time t , if it obeys Hooke's law?
2. Find the average of the force exerted by the spring from $t = 0$ to $t = 6$.
3. Find the root mean square of the force exerted by the spring from $t = 0$ to $t = 6$.

Stage 3

? 19 (*)

A car travels two hours without stopping. The driver records the car's speed every 20 minutes, as indicated in the table below:

time in hours	0	1/3	2/3	1	4/3	5/3	2
speed in km/hr	50	70	80	55	60	80	40

1. Use the trapezoidal rule to estimate the total distance traveled in the two hours.
2. Use the answer to part (a) to estimate the average speed of the car during this period.

? 20

Let $s(t) = e^t$.

1. Find the average of $s(t)$ on the interval $[0, 1]$. Call this quantity A .
2. For any point t , the difference between $s(t)$ and A is $s(t) - A$. Find the average value of $s(t) - A$ on the interval $[0, 1]$.
3. For any point t , the absolute difference between $s(t)$ and A is $|s(t) - A|$. Find the average value of $|s(t) - A|$ on the interval $[0, 1]$.

? 21

Consider the two functions $f(x)$ and $g(x)$ below, both of which have average A on $[0, 4]$.

1. Which function has a larger average on $[0, 4]$: $f(x) - A$ or $g(x) - A$?
2. Which function has a larger average on $[0, 4]$: $|f(x) - A|$ or $|g(x) - A|$?

? 22

Suppose the root mean square of a function $f(x)$ on the interval $[a, b]$ is R . What is the volume of the solid formed by rotating the portion of $f(x)$ from a to b about the x -axis?

As in Example 2.2.6, let the root mean square of $f(x)$ on $[a, b]$ be $\sqrt{\frac{1}{b-a} \int_a^b f^2(x) dx}$.

? 23

Suppose $f(x) = ax^2 + bx + c$, and the average value of $f(x)$ on the interval $[0, 1]$ is the same as the average of $f(0)$ and $f(1)$. What is a ?

? 24

Suppose $f(x) = ax^2 + bx + c$, and the average value of $f(x)$ on the interval $[s, t]$ is the same as the average of $f(s)$ and $f(t)$. Is it possible that $a \neq 0$?

That is — does the result of Question 23 generalize?

? 25

Let $f(x)$ be a function defined for all numbers in the interval $[a, b]$, with average value A over that interval. What is the average of $f(a+b-x)$ over the interval $[a, b]$?

? 26

Suppose $f(t)$ is a continuous function, and $A(x)$ is the average of $f(t)$ on the interval from 0 to x .

1. What is the average of $f(t)$ on $[a, b]$, where $a < b$? Give your answer in terms of A .
2. What is $f(t)$? Again, give your answer in terms of A .

? 27

1. Find a function $f(x)$ with average 0 over $[-1, 1]$ but $f(x) \neq 0$ for all x in $[-1, 1]$, or show that no such function exists.
2. Find a *continuous* function $f(x)$ with average 0 over $[-1, 1]$ but $f(x) \neq 0$ for all x in $[-1, 1]$, or show that no such function exists.

? 28

Suppose $f(x)$ is a positive, continuous function with $\lim_{x \rightarrow \infty} f(x) = 0$, and let $A(x)$ be the average of $f(x)$ on $[0, x]$.

True or false: $\lim_{x \rightarrow \infty} A(x) = 0$.

? 29

Let $A(x)$ be the average of the function $f(t) = e^{-t^2}$ on the interval $[0, x]$. What is $\lim_{x \rightarrow \infty} A(x)$?

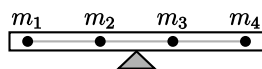
1. Awful pun. The two main approaches to statistics are frequentism and Bayesianism; the latter named after Bayes' Theorem which is, in turn, named for Reverend Thomas Bayes. While this (both the approaches to statistics and their history and naming) is a very interesting and quite philosophical topic, it is beyond the scope of this course. The interested reader has plenty of interesting reading here to interest them.
2. We are being a little loose here with the distinction between mean and average. To be much more pedantic — the average is the arithmetic mean. Other interesting “means” are the geometric and harmonic means:
3. A normal household socket delivers alternating current, rather than the direct current USB supplies. At the risk of yet another “the interested reader” suggestion — the how and why household plugs supply AC current is another worthwhile and interesting digression from studying integration. The interested reader should look up the “War of Currents”. The diligent and interested reader should bookmark this, finish the section and come back to it later.
4. Some countries supply power at 50 cycles per second. Japan actually supplies both — 50 cycles in the east of the country and 60 in the west.
5. This example was written in North America where the standard voltage supplied to homes is 120 volts. Most of the rest of the world supplies homes with 240 volts. The main reason for this difference is the development of the light bulb. The USA electrified earlier when the best voltage for bulb technology was 110 volts. As time went on, bulb technology improved and countries that electrified later took advantage of this (and the cheaper transmission costs that come with higher voltage) and standardised at 240 volts. So many digressions in this section!
6. For a finite set of numbers one can compute the “quadratic mean” which is another way to generalise the notion of the average:
7. A quick glance at Appendix A.14 will refresh your memory.
8. There are many unsurprising things that are true, but there are also many unsurprising things that surprisingly turn out to be false. Mathematicians like to prove things - surprising or not.

This page titled [2.2: Averages](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

2.3: Center of Mass and Torque

Center of Mass

If you support a body at its center of mass (in a uniform gravitational field) it balances perfectly. That's the definition of the center of mass of the body.



If the body consists of a finite number of masses m_1, \dots, m_n attached to an infinitely strong, weightless (idealized) rod with mass number i attached at position x_i , then the center of mass is at the (weighted) average value of x :

Equation 2.3.1 Center of mass (discrete masses)

$$\bar{x} = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i}$$

The denominator $m = \sum_{i=1}^n m_i$ is the total mass of the body.

This formula for the center of mass is derived in the following (optional) section. See equation (2.3.14).

For many (but certainly not all) purposes an (extended rigid) body acts like a point particle located at its center of mass. For example it is very common to treat the Earth as a point particle. Here is a more detailed example in which we think of a body as being made up of a number of component parts and compute the center of mass of the body as a whole by using the center of masses of the component parts. Suppose that we have a dumbbell which consists of

- a left end made up of particles of masses $m_{l,1}, \dots, m_{l,3}$ located at $x_{l,1}, \dots, x_{l,3}$ and
- a right end made up of particles of masses $m_{r,1}, \dots, m_{r,4}$ located at $x_{r,1}, \dots, x_{r,4}$ and
- an infinitely strong, weightless (idealized) rod joining all of the particles.

Then the mass and center of mass of the left end are

$$M_l = m_{l,1} + \dots + m_{l,3} \quad \bar{X}_l = \frac{m_{l,1}x_{l,1} + \dots + m_{l,3}x_{l,3}}{M_l}$$

and the mass and center of mass of the right end are

$$M_r = m_{r,1} + \dots + m_{r,4} \quad \bar{X}_r = \frac{m_{r,1}x_{r,1} + \dots + m_{r,4}x_{r,4}}{M_r}$$

The mass and center of mass of the entire dumbbell are

$$\begin{aligned} M &= m_{l,1} + \dots + m_{l,3} + m_{r,1} + \dots + m_{r,4} \\ &= M_l + M_r \\ \bar{x} &= \frac{m_{l,1}x_{l,1} + \dots + m_{l,3}x_{l,3} + m_{r,1}x_{r,1} + \dots + m_{r,4}x_{r,4}}{M} \\ &= \frac{M_l \bar{X}_l + M_r \bar{X}_r}{M_r + M_l} \end{aligned}$$

So we can compute the center of mass of the entire dumbbell by treating it as being made up of two point particles, one of mass M_l located at the center of mass of the left end, and one of mass M_r located at the center of mass of the right end.

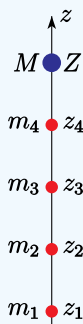
Example 2.3.2 Work and Center of Mass

Here is another example in which an extended body acts like a point particle located at its center of mass. Imagine that there are a finite number of masses m_1, \dots, m_n arrayed along a (vertical) z -axis with mass number i attached at height z_i . Note that the total mass of the array is $M = \sum_{i=1}^n m_i$ and that the center of mass of the array is at height

$$\bar{z} = \frac{\sum_{i=1}^n m_i z_i}{\sum_{i=1}^n m_i} = \frac{1}{M} \sum_{i=1}^n m_i z_i$$

Now suppose that we lift all of the masses, against gravity, to height Z . So after the lift there is a total mass M located at height Z . The i^{th} mass is subject to a downward gravitational force of $m_i g$. So to lift the i^{th} mass we need to apply a compensating upward force of $m_i g$ through a distance of $Z - z_i$. This takes work $m_i g(Z - z_i)$. So the total work required to lift all n masses is

$$\begin{aligned} \text{Work} &= \sum_{i=1}^n m_i g(Z - z_i) \\ &= gZ \sum_{i=1}^n m_i - g \sum_{i=1}^n m_i z_i \\ &= gZM - gM\bar{z} \\ &= Mg(Z - \bar{z}) \end{aligned}$$



So the work required to lift the array of n particles is identical to the work required to lift a single particle, whose mass, M , is the total mass of the array, from height \bar{z} , the center of mass of the array, to height Z .

✓ Example 2.3.3 Example 2.3.2, continued

Imagine, as in Example 2.3.2, that there are a finite number of masses m_1, \dots, m_n arrayed along a (vertical) z -axis with mass number i attached at height z_i . Again, the total mass and center of mass of the array are

$$M = \sum_{i=1}^n m_i \quad \bar{z} = \frac{\sum_{i=1}^n m_i z_i}{\sum_{i=1}^n m_i} = \frac{1}{M} \sum_{i=1}^n m_i z_i$$

Now suppose that we lift, for each $1 \leq i \leq n$, mass number i , against gravity, from its initial height z_i to a final height Z_i . So after the lift we have a new array of masses with total mass and center of mass


$$M = \sum_{i=1}^n m_i \quad \bar{Z} = \frac{\sum_{i=1}^n m_i Z_i}{\sum_{i=1}^n m_i} = \frac{1}{M} \sum_{i=1}^n m_i Z_i$$

To lift the i^{th} mass took work $m_i g(Z_i - z_i)$. So the total work required to lift all n masses was

$$\begin{aligned} \text{Work} &= \sum_{i=1}^n m_i g(Z_i - z_i) \\ &= g \sum_{i=1}^n m_i Z_i - g \sum_{i=1}^n m_i z_i \\ &= gM\bar{Z} - gM\bar{z} = Mg(\bar{Z} - \bar{z}) \end{aligned}$$

So the work required to lift the array of n particles is identical to the work required to lift a single particle, whose mass, M , is the total mass of the array, from height \bar{z} , the initial center of mass of the array, to height \bar{Z} , the final center of mass of the array.

Now we'll extend the above ideas to cover more general classes of bodies. If the body consists of mass distributed continuously along a straight line, say with mass density $\rho(x)$ kg/m and with x running from a to b , rather than consisting of a finite number of point masses, the formula for the center of mass becomes

 Equation 2.3.4 Center of mass (continuous mass)

$$\bar{x} = \frac{\int_a^b x \rho(x) dx}{\int_a^b \rho(x) dx}$$

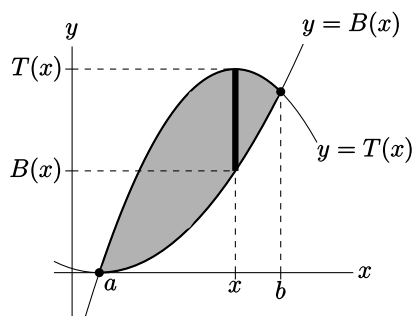
Think of $\rho(x) dx$ as the mass of the “almost point particle” between x and $x + dx$.

If the body is a two dimensional object, like a metal plate, lying in the xy -plane, its center of mass is a point (\bar{x}, \bar{y}) with \bar{x} being the (weighted) average value of the x -coordinate over the body and \bar{y} being the (weighted) average value of the y -coordinate over the body. To be concrete, suppose the body fills the region

$$\{ (x, y) \mid a \leq x \leq b, B(x) \leq y \leq T(x) \}$$

in the xy -plane. For simplicity, we will assume that the density of the body is a constant, say ρ . When the density is constant, the center of mass is also called the *centroid* and is thought of as the geometric center of the body.

To find the centroid of the body, we use our standard “slicing” strategy. We slice the body into thin vertical strips, as illustrated in the figure below.



Here is a detailed description of a generic strip.

- The strip has width dx .
- Each point of the strip has essentially the same x -coordinate. Call it x .
- The top of the strip is at $y = T(x)$ and the bottom of the strip is at $y = B(x)$.
- So the strip has
 - height $T(x) - B(x)$
 - area $[T(x) - B(x)] dx$
 - mass $\rho[T(x) - B(x)] dx$
 - centroid, i.e. middle point, $(x, \frac{B(x)+T(x)}{2})$.

In computing the centroid of the entire body, we may treat each strip as a single particle of mass $\rho[T(x) - B(x)] dx$ located at $(x, \frac{B(x)+T(x)}{2})$. So:

 Equation 2.3.5 Centroid of object with constant density

The mass of the entire body bounded by curves $T(x)$ above and $B(x)$ below is

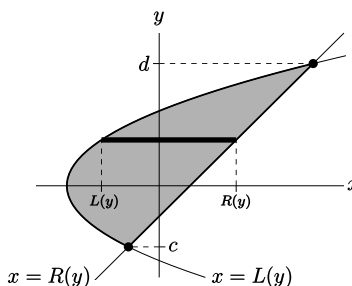
$$M = \rho \int_a^b [T(x) - B(x)] dx = \rho A \tag{a}$$

where $A = \int_a^b [T(x) - B(x)] dx$ is the area of the region. The coordinates of the centroid are

$$\bar{x} = \frac{\int_a^b x \overbrace{\rho[T(x) - B(x)] dx}^{\text{mass of slice}}}{M} = \frac{\int_a^b x [T(x) - B(x)] dx}{A} \quad (b)$$

$$\bar{y} = \frac{\int_a^b \overbrace{\frac{B(x)+T(x)}{2}}^{\text{average } y \text{ on slice}} \overbrace{\rho[T(x) - B(x)] dx}^{\text{mass of slice}}}{M} = \frac{\int_a^b [T(x)^2 - B(x)^2] dx}{2A} \quad (c)$$

We can of course also slice up the body using horizontal slices.



If the body has constant density ρ and fills the region

$$\{ (x, y) \mid L(y) \leq x \leq R(y), c \leq y \leq d \}$$

then the same computation as above gives:

Equation 2.3.6 Centroid of object with constant density

The mass of the entire body bounded by curves $L(y)$ to the left and $R(y)$ to the right is

$$M = \rho \int_c^d [R(y) - L(y)] dy = \rho A \quad (a)$$

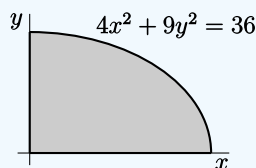
where $A = \int_c^d [R(y) - L(y)] dy$ is the area of the region, and gives the coordinates of the centroid to be

$$\bar{x} = \frac{\int_c^d \overbrace{\frac{R(y)+L(y)}{2}}^{\text{average } x \text{ on slice}} \overbrace{\rho[R(y) - L(y)] dy}^{\text{mass of slice}}}{M} = \frac{\int_c^d [R(y)^2 - L(y)^2] dy}{2A} \quad (b)$$

$$\bar{y} = \frac{\int_c^d \overbrace{y \rho[R(y) - L(y)] dy}^{\text{mass of slice}}}{M} = \frac{\int_c^d y [R(y) - L(y)] dy}{A} \quad (c)$$

Example 2.3.7 Centroid of a quarter ellipse

Find the x -coordinate of the centroid (center of gravity) of the plane region R that lies in the first quadrant $x \geq 0, y \geq 0$ and inside the ellipse $4x^2 + 9y^2 = 36$. (The area bounded by the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ is πab square units.)



Solution: In standard form $4x^2 + 9y^2 = 36$ is $\frac{x^2}{9} + \frac{y^2}{4} = 1$. So, on R , x runs from 0 to 3 and R has area $A = \frac{1}{4} \pi \times 3 \times 2 = \frac{3}{2} \pi$. For each fixed x , between 0 and 3, y runs from 0 to $2\sqrt{1 - \frac{x^2}{9}}$. So, applying (2.3.5.b) with $a = 0$, $b = 3$, $T(x) = 2\sqrt{1 - \frac{x^2}{9}}$ and $B(x) = 0$,

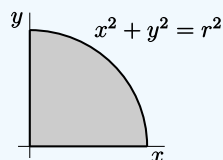
$$\bar{x} = \frac{1}{A} \int_0^3 x T(x) dx = \frac{1}{A} \int_0^3 x 2\sqrt{1 - \frac{x^2}{9}} dx = \frac{4}{3\pi} \int_0^3 x \sqrt{1 - \frac{x^2}{9}} dx$$

Sub in $u = 1 - \frac{x^2}{9}$, $du = -\frac{2}{9}x dx$.

$$\bar{x} = -\frac{9}{2} \frac{4}{3\pi} \int_1^0 \sqrt{u} du = -\frac{9}{2} \frac{4}{3\pi} \left[\frac{u^{3/2}}{3/2} \right]_1^0 = -\frac{9}{2} \frac{4}{3\pi} \left[-\frac{2}{3} \right] = \frac{4}{\pi}$$

✓ Example 2.3.8 Centroid of a quarter disk

Find the centroid of the quarter circular disk $x \geq 0$, $y \geq 0$, $x^2 + y^2 \leq r^2$.



Solution: By symmetry, $\bar{x} = \bar{y}$. The area of the quarter disk is $A = \frac{1}{4}\pi r^2$. By (2.3.5.b) with $a = 0$, $b = r$, $T(x) = \sqrt{r^2 - x^2}$ and $B(x) = 0$,

$$\bar{x} = \frac{1}{A} \int_0^r x \sqrt{r^2 - x^2} dx$$

To evaluate the integral, sub in $u = r^2 - x^2$, $du = -2x dx$.

$$\int_0^r x \sqrt{r^2 - x^2} dx = \int_{r^2}^0 \sqrt{u} \frac{du}{-2} = -\frac{1}{2} \left[\frac{u^{3/2}}{3/2} \right]_{r^2}^0 = \frac{r^3}{3}$$

So

$$\bar{x} = \frac{4}{\pi r^2} \left[\frac{r^3}{3} \right] = \frac{4r}{3\pi}$$

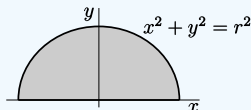
As we observed above, we should have $\bar{x} = \bar{y}$. But, just for practice, let's compute \bar{y} by the integral formula (2.3.5.c), again with $a = 0$, $b = r$, $T(x) = \sqrt{r^2 - x^2}$ and $B(x) = 0$,

$$\begin{aligned} \bar{y} &= \frac{1}{2A} \int_0^r (\sqrt{r^2 - x^2})^2 dx &= \frac{2}{\pi r^2} \int_0^r (r^2 - x^2) dx \\ &= \frac{2}{\pi r^2} \left[r^2 x - \frac{x^3}{3} \right]_0^r &= \frac{2}{\pi r^2} \frac{2r^3}{3} \\ &= \frac{4r}{3\pi} \end{aligned}$$

as expected.

✓ Example 2.3.9 Centroid of a half disc

Find the centroid of the half circular disk $y \geq 0$, $x^2 + y^2 \leq r^2$.

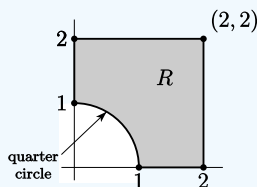


Solution: Once again, we have a symmetry — namely the half disk is symmetric about the y -axis. So the centroid lies on the y -axis and $\bar{x} = 0$. The area of the half disk is $A = \frac{1}{2}\pi r^2$. By (2.3.5.c), with $a = -r$, $b = r$, $T(x) = \sqrt{r^2 - x^2}$ and $B(x) = 0$,

$$\begin{aligned}\bar{y} &= \frac{1}{2A} \int_{-r}^r (\sqrt{r^2 - x^2})^2 dx &&= \frac{1}{\pi r^2} \int_{-r}^r (r^2 - x^2) dx \\ &= \frac{2}{\pi r^2} \int_0^r (r^2 - x^2) dx &&\text{since the integrand is even} \\ &= \frac{2}{\pi r^2} \left[r^2 x - \frac{x^3}{3} \right]_0^r \\ &= \frac{4r}{3\pi}\end{aligned}$$

✓ Example 2.3.10 Another centroid

Find the centroid of the region R in the diagram.



Solution: By symmetry, $\bar{x} = \bar{y}$. The region R is a 2×2 square with one quarter of a circle of radius 1 removed and so has area $2 \times 2 - \frac{1}{4}\pi = \frac{16-\pi}{4}$. The top of R is $y = T(x) = 2$. The bottom is $y = B(x)$ with $B(x) = \sqrt{1-x^2}$ when $0 \leq x \leq 1$ and $B(x) = 0$ when $1 \leq x \leq 2$. So

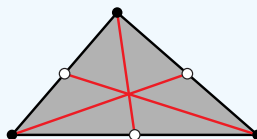
$$\begin{aligned}\bar{y} = \bar{x} &= \frac{1}{A} \left[\int_0^1 x[2 - \sqrt{1-x^2}] dx + \int_1^2 x[2-0] dx \right] \\ &= \frac{4}{16-\pi} \left[x^2 \Big|_0^1 + x^2 \Big|_1^2 - \int_0^1 x\sqrt{1-x^2} dx \right]\end{aligned}$$

Now we can make use of the starred equation in Example 2.3.8 with $r = 1$ to obtain

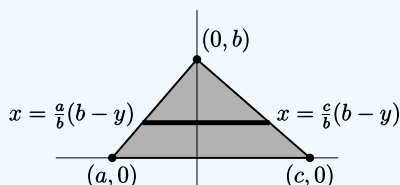
$$\begin{aligned}&= \frac{4}{16-\pi} \left[4 - \frac{1}{3} \right] \\ &= \frac{44}{48-3\pi}\end{aligned}$$

✓ Example 2.3.11 Centroid of a triangle and its medians

Prove that the centroid of any triangle is located at the point of intersection of the medians. A median of a triangle is a line segment joining a vertex to the midpoint of the opposite side.



Solution: Choose a coordinate system so that the vertices of the triangle are located at $(a, 0)$, $(0, b)$ and $(c, 0)$. (In the figure below, a is negative.)



The line joining $(a, 0)$ and $(0, b)$ has equation $bx + ay = ab$. (Check that $(a, 0)$ and $(0, b)$ both really are on this line.) The line joining $(c, 0)$ and $(0, b)$ has equation $bx + cy = bc$. (Check that $(c, 0)$ and $(0, b)$ both really are on this line.) Hence for each fixed y between 0 and b , x runs from $a - \frac{a}{b}y$ to $c - \frac{c}{b}y$.

We'll use horizontal strips to compute \bar{x} and \bar{y} . We could just apply equation (2.3.6) with $c = 0$, $d = b$, $R(y) = \frac{c}{b}(b - y)$ (which is gotten by solving $bx + cy = bc$ for x) and $L(y) = \frac{a}{b}(b - y)$ (which is gotten by solving $bx + ay = ab$ for x).

But rather than memorizing or looking up those formulae, we'll derive them for this example. So consider a thin strip at height y as illustrated in the figure above.

- The strip has length

$$\ell(y) = \left[\frac{c}{b}(b - y) - \frac{a}{b}(b - y) \right] = \frac{c - a}{b}(b - y)$$

- The strip has width dy .
- On this strip, y has average value y .
- On this strip, x has average value $\frac{1}{2} \left[\frac{a}{b}(b - y) + \frac{c}{b}(b - y) \right] = \frac{a + c}{2b}(b - y)$.

As the area of the triangle is $A = \frac{1}{2}(c - a)b$,

$$\begin{aligned} \bar{y} &= \frac{1}{A} \int_0^b y \ell(y) dy = \frac{2}{(c - a)b} \int_0^b y \frac{c - a}{b}(b - y) dy \\ &= \frac{2}{b^2} \int_0^b (by - y^2) dy = \frac{2}{b^2} \left(b \frac{b^2}{2} - \frac{b^3}{3} \right) \\ &= \frac{2}{b^2} \frac{b^3}{6} = \frac{b}{3} \\ \bar{x} &= \frac{1}{A} \int_0^b \frac{a + c}{2b}(b - y) \ell(y) dy \\ &= \frac{2}{(c - a)b} \int_0^b \frac{a + c}{2b}(b - y) \frac{c - a}{b}(b - y) dy \\ &= \frac{a + c}{b^3} \int_0^b (y - b)^2 dy \\ &= \frac{a + c}{b^3} \left[\frac{1}{3}(y - b)^3 \right]_0^b = \frac{a + c}{b^3} \frac{b^3}{3} = \frac{a + c}{3} \end{aligned}$$

We have found that the centroid of the triangle is at $(\bar{x}, \bar{y}) = \left(\frac{a + c}{3}, \frac{b}{3} \right)$. We shall now show that this point lies on all three medians.

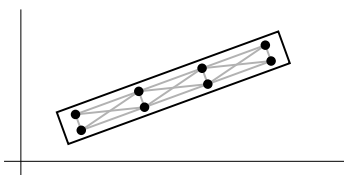
- One vertex is at $(a, 0)$. The opposite side runs from $(0, b)$ and $(c, 0)$ and so has midpoint $\frac{1}{2}(c, b)$. The line from $(a, 0)$ to $\frac{1}{2}(c, b)$ has slope $\frac{b/2}{c/2 - a} = \frac{b}{c - 2a}$ and so has equation $y = \frac{b}{c - 2a}(x - a)$. As $\frac{b}{c - 2a}(\bar{x} - a) = \frac{b}{c - 2a} \left(\frac{a + c}{3} - a \right) = \frac{1}{3} \frac{b}{c - 2a}(c + a - 3a) = \frac{b}{3} = \bar{y}$, the centroid does indeed lie on this median. In this computation we have implicitly assumed that $c \neq 2a$ so that the denominator $c - 2a \neq 0$. In the event that $c = 2a$, the median runs from $(a, 0)$ to $(a, \frac{b}{2})$ and so has equation $x = a$. When $c = 2a$ we also have $\bar{x} = \frac{a + c}{3} = a$, so that the centroid still lies on the median.
- Another vertex is at $(c, 0)$. The opposite side runs from $(a, 0)$ and $(0, b)$ and so has midpoint $\frac{1}{2}(a, b)$. The line from $(c, 0)$ to $\frac{1}{2}(a, b)$ has slope $\frac{b/2}{a/2 - c} = \frac{b}{a - 2c}$ and so has equation $y = \frac{b}{a - 2c}(x - c)$. As $\frac{b}{a - 2c}(\bar{x} - c) = \frac{b}{a - 2c} \left(\frac{a + c}{3} - c \right) = \frac{1}{3} \frac{b}{a - 2c}(a + c - 3c) = \frac{b}{3} = \bar{y}$, the centroid does indeed lie on this median. In this computation we have implicitly assumed that $a \neq 2c$ so that the denominator $a - 2c \neq 0$. In the event that $a = 2c$, the median runs from $(c, 0)$ to $(c, \frac{b}{2})$ and so has equation $x = c$. When $a = 2c$ we also have $\bar{x} = \frac{a + c}{3} = c$, so that the centroid still lies on the median.
- The third vertex is at $(0, b)$. The opposite side runs from $(a, 0)$ and $(c, 0)$ and so has midpoint $\left(\frac{a + c}{2}, 0 \right)$. The line from $(0, b)$ to $\left(\frac{a + c}{2}, 0 \right)$ has slope $\frac{-b}{(a + c)/2} = -\frac{2b}{a + c}$ and so has equation $y = b - \frac{2b}{a + c}x$. As $b - \frac{2b}{a + c}\bar{x} = b - \frac{2b}{a + c} \frac{a + c}{3} = \frac{b}{3} = \bar{y}$, the centroid does indeed lie on this median. This time, we have implicitly assumed that $a + c \neq 0$. In the event that

$a + c = 0$, the median runs from $(0, b)$ to $(0, 0)$ and so has equation $x = 0$. When $a + c = 0$ we also have $\bar{x} = \frac{a+c}{3} = 0$, so that the centroid still lies on the median.

Optional — Torque

Newton's law of motion says that the position $x(t)$ of a single particle moving under the influence of a force F obeys $m x''(t) = F$. Similarly, the positions $x_i(t)$, $1 \leq i \leq n$, of a set of particles moving under the influence of forces F_i obey $m x_i''(t) = F_i$, $1 \leq i \leq n$. Often systems of interest consist of some small number of rigid bodies. Suppose that we are interested in the motion of a single rigid body, say a piece of wood. The piece of wood is made up of a huge number of atoms. So the system of equations determining the motion of all of the individual atoms in the piece of wood is huge. On the other hand, because the piece of wood is rigid, its configuration is completely determined by the position of, for example, its center of mass and its orientation. (Rather than get into what is precisely meant by "orientation", let's just say that it is certainly determined by, for example, the positions of a few of the corners of the piece of wood). It is possible to extract from the huge system of equations that determine the motion of all of the individual atoms, a small system of equations that determine the motion of the center of mass and the orientation. We can avoid some vector analysis, that is beyond the scope of this course, by assuming that our rigid body is moving in two rather than three dimensions.

So, imagine a piece of wood moving in the xy -plane.



Furthermore, imagine that the piece of wood consists of a huge number of particles joined by a huge number of weightless but very strong steel rods. The steel rod joining particle number one to particle number two just represents a force acting between particles number one and two. Suppose that

- there are n particles, with particle number i having mass m_i
- at time t , particle number i has x -coordinate $x_i(t)$ and y -coordinate $y_i(t)$
- at time t , the external force (gravity and the like) acting on particle number i has x -coordinate $H_i(t)$ and y -coordinate $V_i(t)$. Here H stands for horizontal and V stands for vertical.
- at time t , the force acting on particle number i , due to the steel rod joining particle number i to particle number j has x -coordinate $H_{i,j}(t)$ and y -coordinate $V_{i,j}(t)$. If there is no steel rod joining particles number i and j , just set $H_{i,j}(t) = V_{i,j}(t) = 0$. In particular, $H_{i,i}(t) = V_{i,i}(t) = 0$.

The only assumptions that we shall make about the steel rod forces are

(A1)

for each $i \neq j$, $H_{i,j}(t) = -H_{j,i}(t)$ and $V_{i,j}(t) = -V_{j,i}(t)$. In words, the steel rod joining particles i and j applies equal and opposite forces to particles i and j .

(A2)

for each $i \neq j$, there is a function $M_{i,j}(t)$ such that $H_{i,j}(t) = M_{i,j}(t)[x_i(t) - x_j(t)]$ and $V_{i,j}(t) = M_{i,j}(t)[y_i(t) - y_j(t)]$. In words, the force due to the rod joining particles i and j acts parallel to the line joining particles i and j . For (A1) to be true, we need $M_{i,j}(t) = M_{j,i}(t)$.

Newton's law of motion, applied to particle number i , now tells us that

$$m_i x_i''(t) = H_i(t) + \sum_{j=1}^n H_{i,j}(t) \quad (X_i)$$

$$m_i y_i''(t) = V_i(t) + \sum_{j=1}^n V_{i,j}(t) \quad (Y_i)$$

Adding up all of the equations (X_i) , for $i = 1, 2, 3, \dots, n$ and adding up all of the equations (Y_i) , for $i = 1, 2, 3, \dots, n$ gives

$$\sum_{i=1}^n m_i x_i''(t) = \sum_{i=1}^n H_i(t) + \sum_{1 \leq i, j \leq n} H_{i,j}(t) \quad (\Sigma_i X_i)$$

$$\sum_{i=1}^n m_i y_i''(t) = \sum_{i=1}^n V_i(t) + \sum_{1 \leq i, j \leq n} V_{i,j}(t) \quad (\Sigma_i Y_i)$$

The sum $\sum_{1 \leq i, j \leq n} H_{i,j}(t)$ contains $H_{1,2}(t)$ exactly once and it also contains $H_{2,1}(t)$ exactly once and these two terms cancel exactly, by assumption (A1). In this way, all terms in $\sum_{1 \leq i, j \leq n} H_{i,j}(t)$ with $i \neq j$ exactly cancel. All terms with $i = j$ are assumed to be zero. So $\sum_{1 \leq i, j \leq n} H_{i,j}(t) = 0$. Similarly, $\sum_{1 \leq i, j \leq n} V_{i,j}(t) = 0$, so the equations $(\Sigma_i X_i)$ and $(\Sigma_i Y_i)$ simplify to

$$\sum_{i=1}^n m_i x_i''(t) = \sum_{i=1}^n H_i(t) \quad (\Sigma_i X_i)$$

$$\sum_{i=1}^n m_i y_i''(t) = \sum_{i=1}^n V_i(t) \quad (\Sigma_i Y_i)$$

Denote by

$$M = \sum_{i=1}^n m_i$$

the total mass of the system, by

$$X(t) = \frac{1}{M} \sum_{i=1}^n m_i x_i(t) \quad \text{and} \quad Y(t) = \frac{1}{M} \sum_{i=1}^n m_i y_i(t)$$

the x - and y -coordinates of the center of mass of the system at time t and by

$$H(t) = \sum_{i=1}^n H_i(t) \quad \text{and} \quad V(t) = \sum_{i=1}^n V_i(t)$$

the x - and y -coordinates of the total external force acting on the system at time t . In this notation, the equations $(\Sigma_i X_i)$ and $(\Sigma_i Y_i)$ are

Equation 2.3.12 Rectilinear motion of center of mass

$$MX''(t) = H(t) \quad MY''(t) = V(t)$$

So the center of mass of the system moves just like a single particle of mass M subject to the total external force.

Now multiply equation (Y_i) by $x_i(t)$, subtract from it equation (X_i) multiplied by $y_i(t)$, and sum over i . This gives the equation $\sum_i [x_i(t)(Y_i) - y_i(t)(X_i)]$:

$$\begin{aligned} \sum_{i=1}^n m_i [x_i(t)y_i''(t) - y_i(t)x_i''(t)] &= \sum_{i=1}^n [x_i(t)V_i(t) - y_i(t)H_i(t)] \\ &\quad + \sum_{1 \leq i, j \leq n} [x_i(t)V_{i,j}(t) - y_i(t)H_{i,j}(t)] \end{aligned}$$

By the assumption (A2)

$$\begin{aligned} x_1(t)V_{1,2}(t) - y_1(t)H_{1,2}(t) &= x_1(t)M_{1,2}(t)[y_1(t) - y_2(t)] - y_1(t)M_{1,2}(t)[x_1(t) - x_2(t)] \\ &= M_{1,2}(t)[y_1(t)x_2(t) - x_1(t)y_2(t)] \\ x_2(t)V_{2,1}(t) - y_2(t)H_{2,1}(t) &= x_2(t)M_{2,1}(t)[y_2(t) - y_1(t)] - y_2(t)M_{2,1}(t)[x_2(t) - x_1(t)] \\ &= M_{2,1}(t)[-y_1(t)x_2(t) + x_1(t)y_2(t)] \\ &= M_{1,2}(t)[-y_1(t)x_2(t) + x_1(t)y_2(t)] \end{aligned}$$

So the $i = 1, j = 2$ term in $\sum_{1 \leq i, j \leq n} [x_i(t)V_{i,j}(t) - y_i(t)H_{i,j}(t)]$ exactly cancels the $i = 2, j = 1$ term. In this way all of the terms in $\sum_{1 \leq i, j \leq n} [x_i(t)V_{i,j}(t) - y_i(t)H_{i,j}(t)]$ with $i \neq j$ cancel. Each term with $i = j$ is exactly zero. So $\sum_{1 \leq i, j \leq n} [x_i(t)V_{i,j}(t) - y_i(t)H_{i,j}(t)] = 0$ and

$$\sum_{i=1}^n m_i [x_i(t)y_i''(t) - y_i(t)x_i''(t)] = \sum_{i=1}^n [x_i(t)V_i(t) - y_i(t)H_i(t)]$$

Define

$$L(t) = \sum_{i=1}^n m_i [x_i(t)y_i'(t) - y_i(t)x_i'(t)]$$

$$T(t) = \sum_{i=1}^n [x_i(t)V_i(t) - y_i(t)H_i(t)]$$

In this notation

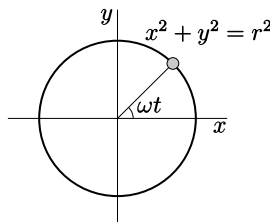
Equation 2.3.13 Rotational motion of center of mass

$$\frac{d}{dt} L(t) = T(t)$$

- Equation (2.3.13) plays the role of Newton's law of motion for rotational motion.
- $T(t)$ is called the torque and plays the role of “rotational force”.
- $L(t)$ is called the angular momentum (about the origin) and is a measure of the rate at which the piece of wood is rotating.
 - For example, if a particle of mass m is traveling in a circle of radius r , centered on the origin, at ω radians per unit time, then $x(t) = r \cos(\omega t)$, $y(t) = r \sin(\omega t)$ and

$$m [x(t)y'(t) - y(t)x'(t)] = m [r \cos(\omega t) r \omega \cos(\omega t) - r \sin(\omega t) (-r \omega \sin(\omega t))] = mr^2 \omega$$

is proportional to ω , which is the rate of rotation about the origin.



In any event, in order for the piece of wood to remain stationary, that is to have $x_i(t)$ and $y_i(t)$ be constant for all $1 \leq i \leq n$, we need to have

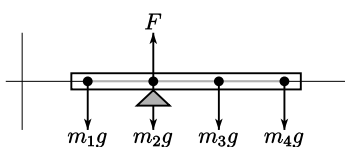
$$X''(y) = Y''(t) = L(t) = 0$$

and then equations (2.3.12) and (2.3.13) force

$$H(t) = V(t) = T(t) = 0$$

Now suppose that the piece of wood is a seesaw that is long and thin and is lying on the x -axis, supported on a fulcrum at $x = p$. Then every $y_i = 0$ and the torque simplifies to $T(t) = \sum_{i=1}^n x_i(t)V_i(t)$. The forces consist of

- gravity, $m_i g$, acting downwards on particle number i , for each $1 \leq i \leq n$ and the
- force F imposed by the fulcrum that is pushing straight up on the particle at $x = p$.



So

- The net vertical force is $V(t) = F - \sum_{i=1}^n m_i g = F - Mg$. If the seesaw is to remain stationary, this must be zero so that $F = Mg$.
- The total torque (about the origin) is

$$T = Fp - \sum_{i=1}^n m_i g x_i = Mgp - \sum_{i=1}^n m_i g x_i$$

If the seesaw is to remain stationary, this must also be zero and the fulcrum must be placed at

Equation 2.3.14 Placement of fulcrum

$$p = \frac{1}{M} \sum_{i=1}^n m_i x_i$$

which is the center of mass of the piece of wood.

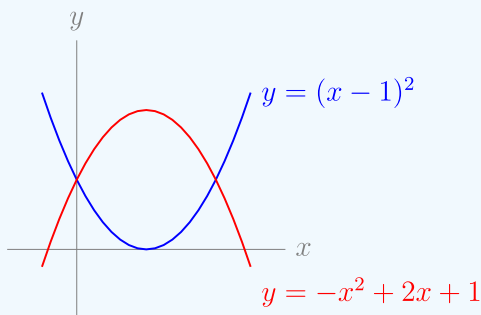
Exercises

Stage 1

In Questions 8 through 10, you will derive the formulas for the center of mass of a rod of variable density, and the centroid of a two-dimensional region using vertical slices (Equations 2.3.4 and 2.3.5 in the text). Knowing the equations by heart will allow you to answer many questions in this section; understanding where they came from will allow you to generalize their ideas to answer even more questions.

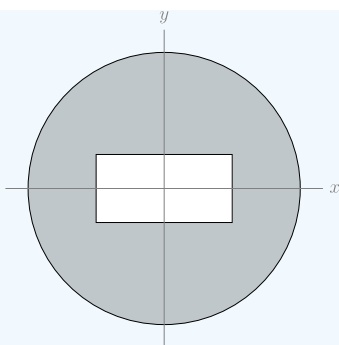
? 1

Using symmetry, find the centroid of the finite region between the curves $y = (x - 1)^2$ and $y = -x^2 + 2x + 1$.



? 2

Using symmetry, find the centroid of the region inside the unit circle, and outside a rectangle centered at the origin with width 1 and height 0.5.



? 3

A long, straight, thin rod has a number of weights attached along it. True or false: if it balances at position x , then the mass to the right of x is the same as the mass to the left of x .

? 4

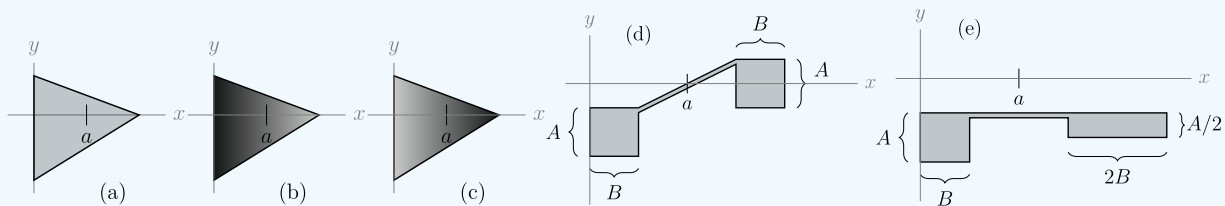
A straight rod with negligible mass has the following weights attached to it:

- A weight of mass 1 kg, 1m from the left end,
- a weight of mass 2 kg, 3m from the left end,
- a weight of mass 2 kg, 4m from the left end, and
- a weight of mass 1 kg, 6m from the left end.

Where is the center of mass of the weighted rod?

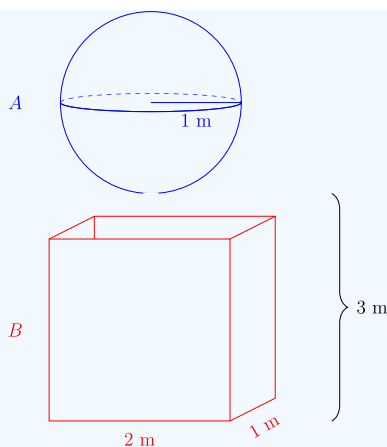
? 5

For each picture below, determine whether the centroid is to the left of, to the right of, or along the line $x = a$, or whether there is not enough information to tell. The shading of a region indicates density: darker shading corresponds to a denser area.



? 6

Tank A is spherical, of radius 1 metre, and filled completely with water. The bottom of tank A is three metres above the ground, where Tank B sits. Tank B is tall and rectangular, with base dimensions 2 metres by 1 metre, and empty. Calculate the work done by gravity to drain all the water from Tank A to Tank B by modelling the situation as a point mass, of the same mass as the water, being moved from the height of the center of mass of A to the height of the center of mass of B .



You may use 1000 kg/m^3 for the density of water, and $g = 9.8 \text{ m/sec}^2$ for the acceleration due to gravity.

? 7

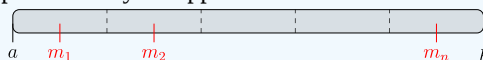
Let S be the region bounded above by $y = \frac{1}{x}$ and below by the x -axis, $1 \leq x \leq 3$. Let R be a rod with density $\rho(x) = \frac{1}{x}$ at position x , $1 \leq x \leq 3$.

1. What is the area of a thin slice of S at position x with width dx ?
2. What is the mass of a small piece of R at position x with length dx ?
3. What is the total area of S ?
4. What is the total mass of R ?
5. What is the x -coordinate of the centroid of S ?
6. What is the center of mass of R ?

? 8

Suppose R is a straight, thin rod with density $\rho(x)$ at a position x . Let the left endpoint of R lie at $x = a$, and the right endpoint lie at $x = b$.

1. To approximate the center of mass of R , imagine chopping it into n pieces of equal length, and approximating the mass of each piece using the density at its midpoint. Give your approximation for the center of mass in sigma notation.



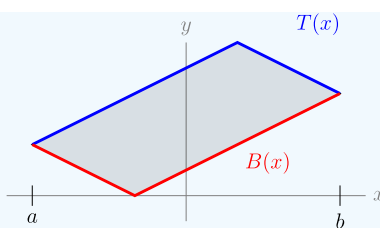
2. Take the limit as n goes to infinity of your approximation in part (a), and express the result using a definite integral.

? 9

Suppose S is a two-dimensional object and at (horizontal) position x its height is $T(x) - B(x)$. Its leftmost point is at position $x = a$, and its rightmost point is at position $x = b$.

To approximate the x -coordinate of the centroid of S , we imagine it as a straight, thin rod R , where the mass of R from $a \leq x \leq b$ is equal to the area of S from $a \leq x \leq b$.

1. If S is the sheet shown below, sketch R as a rod with the same horizontal length, shaded darker when R is denser, and lighter when R is less dense.



2. If we cut S into strips of very small width dx , what is the area of the strip at position x ?
3. Using your answer from (b), what is the density $\rho(x)$ of R at position x ?
4. Using your result from Question 8(b), give the x -coordinate of the centroid of S . Your answer will be in terms of a , b , $T(x)$, and $B(x)$.

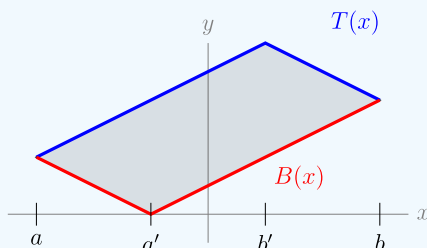
? 10

Suppose S is flat sheet with uniform density, and at (horizontal) position x its height is $T(x) - B(x)$. Its leftmost point is at position $x = a$, and its rightmost point is at position $x = b$.

To approximate the y -coordinate of the centroid of S , we imagine it as a straight, thin, vertical rod R . We slice S into thin, vertical strips, and model these as weights on R with:

- position y on R , where y is the center of mass of the strip, and
- mass in R equal to the area of the strip in S .

1. If S is the sheet shown below, slice it into a number of vertical pieces of equal length, approximated by rectangles. For each rectangle, mark its center of mass. Sketch R as a rod with the same vertical height, with weights corresponding to the slices you made of S .



2. Imagine a thin strip of S at position x , with thickness dx . What is the area of the strip? What is the y -value of its center of mass?
3. Recall the center of mass of a rod with n weights of mass M_i at position y_i is given by

$$\frac{\sum_{i=1}^n (M_i \times y_i)}{\sum_{i=1}^n M_i}$$

Considering the limit of this formula as n goes to infinity, give the y -coordinate of the center of mass of S .

? 11 (*)

Express the x -coordinate of the centroid of the triangle with vertices $(-1, -3)$, $(-1, 3)$, and $(0, 0)$ in terms of a definite integral. **Do not evaluate the integral.**

Stage 2

Use Equations 2.3.4 and 2.3.5 to find centroids and centers of mass in Questions 12 through 23.

? 12

A long, thin rod extends from $x = 0$ to $x = 7$ metres, and its density at position x is given by $\rho(x) = x$ kg/m. Where is the center of mass of the rod?

? 13

A long, thin rod extends from $x = -3$ to $x = 10$ metres, and its density at position x is given by $\rho(x) = \frac{1}{1+x^2}$ kg/m. Where is the center of mass of the rod?

? 14 (*)

Find the y -coordinate of the centroid of the region bounded by the curves $y = 1$, $y = -e^x$, $x = 0$ and $x = 1$. You may use the fact that the area of this region equals e .

? 15 (*)

Consider the region bounded by $y = \frac{1}{\sqrt{16-x^2}}$, $y = 0$, $x = 0$ and $x = 2$.

1. Sketch this region.
2. Find the y -coordinate of the centroid of this region.

? 16 (*)

Find the centroid of the finite region bounded by $y = \sin(x)$, $y = \cos(x)$, $x = 0$, and $x = \pi/4$.

? 17 (*)

Let A denote the area of the plane region bounded by $x = 0$, $x = 1$, $y = 0$ and $y = \frac{k}{\sqrt{1+x^2}}$, where k is a positive constant.

1. Find the coordinates of the centroid of this region in terms of k and A .
2. For what value of k is the centroid on the line $y = x$?

? 18 (*)

The region R is the portion of the plane which is above the curve $y = x^2 - 3x$ and below the curve $y = x - x^2$.

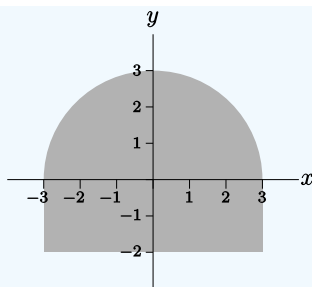
1. Sketch the region R
2. Find the area of R .
3. Find the x coordinate of the centroid of R .

? 19 (*)

Let R be the region where $0 \leq x \leq 1$ and $0 \leq y \leq \frac{1}{1+x^2}$. Find the x -coordinate of the centroid of R .

? 20 (*)

Find the centroid of the region below, which consists of a semicircle of radius 3 on top of a rectangle of width 6 and height 2.



? 21 (*)

Let D be the region below the graph of the curve $y = \sqrt{9 - 4x^2}$ and above the x -axis.

1. Using an appropriate integral, find the *area* of the region D ; simplify your answer completely.
2. Find the *center of mass* of the region D ; simplify your answer completely. (Assume it has constant density ρ .)

? 22

The finite region S is bounded by the lines $y = \arcsin x$, $y = \arcsin(2 - x)$, and $y = -\frac{\pi}{2}$. Find the centroid of S .

? 23

Calculate the centroid of the figure bounded by the curves $y = e^x$, $y = 3(x - 1)$, $y = 0$, $x = 0$, and $x = 2$.

Stage 3

? 24 (*)

Find the y -coordinate of the center of mass of the (infinite) region lying to the right of the line $x = 1$, above the x -axis, and below the graph of $y = 8/x^3$.

? 25 (*)

Let A be the region to the right of the y -axis that is bounded by the graphs of $y = x^2$ and $y = 6 - x$.

1. Find the centroid of A , assuming it has constant density $\rho = 1$. The area of A is $\frac{22}{3}$ (you don't have to show this).
2. Write down an expression, using horizontal slices (disks), for the volume obtained when the region A is rotated around the y -axis. Do not evaluate any integrals; simply write down an expression for the volume.

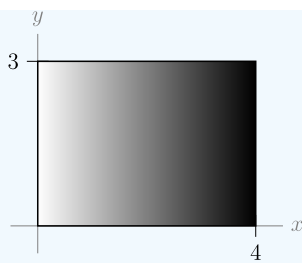
? 26 (*)

(a) Find the y -coordinate of the centroid of the region bounded by $y = e^x$, $x = 0$, $x = 1$, and $y = -1$.

(b) Calculate the volume of the solid generated by rotating the region from part (a) about the line $y = -1$.

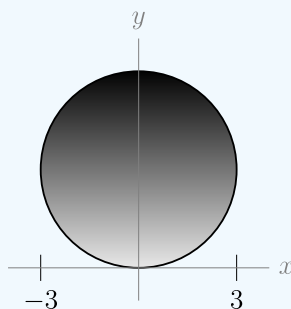
? 27

Suppose a rectangle has width 4 m, height 3 m, and its density x metres from its left edge is x^2 kg/m². Find the center of mass of the rectangle.



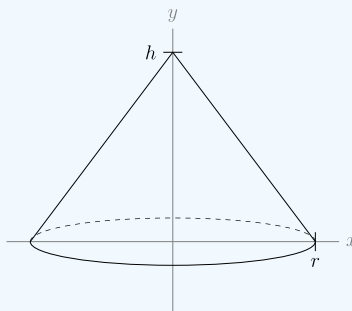
? 28

Suppose a circle of radius 3 m has density $(2 + y)$ kg/m² at any point y metres above its bottom. Find the center of mass of the circle.

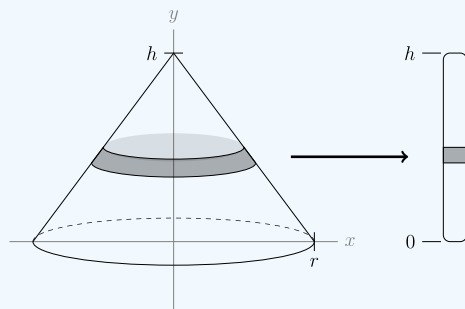


? 29

A right circular cone of uniform density has base radius r m and height h m. We want to find its center of mass. By symmetry, we know that the center of mass will occur somewhere along the straight vertical line through the tip of the cone and the center of its base. The only question is the *height* of the center of mass.



We will model the cone as a rod R with height h , such that the mass of the section of the rod from position a to position b is the same as the volume of the cone from height a to height b . (You can imagine that the cone is an umbrella, and we've closed it up to look like a cane.²)

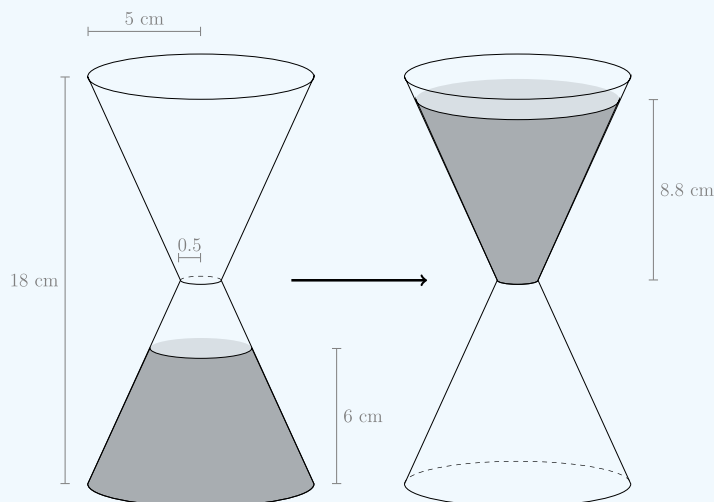


- Using this model, calculate how high above the base of the cone its center of mass is.

2. If we cut off the top $h - k$ metres of the cone (leaving an object of height k), how high above the base is the new center of mass?

? 30

An hourglass is shaped like two identical truncated cones attached together. Their base radius is 5 cm, the height of the entire hourglass is 18 cm, and the radius at the thinnest point is .5 cm. The hourglass contains sand that fills up the bottom 6 cm when it's settled, with mass 600 grams and uniform density. We want to know the work done flipping the hourglass smoothly, so the sand settles into a truncated, inverted-cone shape before it starts to fall down.

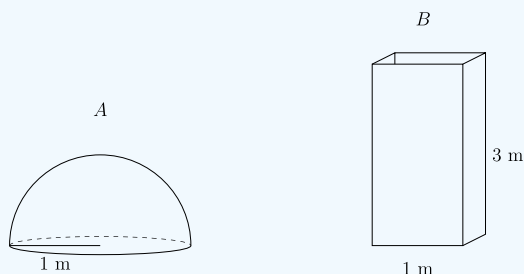


Using the methods of Section 2.1 to calculate the work done would be quite tedious. Instead, we will model the sand as a point of mass 0.6 kg, being lifted from the center of mass of its original position to the center of mass of its upturned position. Using the results of Question 29, how much work was done on the sand?

To simplify your calculation, you may assume that the height of the upturned sand (that is, the distance from the skinniest part of the hourglass to the top of the sand) is 8.8 cm. (Actually, it's $\sqrt[3]{937} - 1 \approx 8.7854$ cm.) So, the top 0.2 cm of the hourglass is empty.

? 31

Tank A is in the shape of half a sphere of radius 1 metre, with its flat face resting on the ground, and is completely filled with water. Tank B is empty and rectangular, with a square base of side length 1 m and a height of 3 m.



1. To pump the water from Tank A to Tank B , we need to pump all the water from Tank A to a height of 3 m. How much work is done to pump all the water from Tank A to a height of 3 m? You may model the water as a point mass, originally situated at the center of mass of the full Tank A .
2. Suppose we could move the water from Tank A directly to its final position in Tank B without going over the top of Tank B . (For example, maybe tank A is elastic, and Tank B is just Tank A after being smooshed into a different form.) How much work is done pumping the water? (That is, how much work is done moving a point mass from the center of mass of Tank A to the center of mass of Tank B ?)

3. What percentage of work from part (a) was “wasted” by pumping the water over the top of Tank B , instead of moving it directly to its final position?

You may assume that the only work done is against the acceleration due to gravity, $g = 9.8 \text{ m/sec}^2$, and that the density of water is 1000 kg/m^3 .

Remark: the answer from (b) is what you might think of as the *net* work involved in pumping the water from Tank A to Tank B . When work gets “wasted,” the pump does some work pumping water up, then gravity does equal and opposite work bringing the water back down.

? 32

Let R be the region bounded above by $y = 2x \sin(x^2)$ and below by the x -axis, $0 \leq x \leq \sqrt{\frac{\pi}{2}}$. Give an approximation of the x -value of the centroid of R with error no more than $\frac{1}{100}$.

You may assume without proof that $\left| \frac{d^4}{dx^4} \{2x^2 \sin(x^2)\} \right| \leq 415$ over the interval $\left[0, \sqrt{\frac{\pi}{2}}\right]$.

This page titled [2.3: Center of Mass and Torque](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

2.4: Separable Differential Equations

A differential equation is an equation for an unknown function that involves the derivative of the unknown function. Differential equations play a central role in modelling a huge number of different phenomena. Here is a table giving a bunch of named differential equations and what they are used for. It is far from complete.

Newton's Law of Motion	describes motion of particles
Maxwell's equations	describes electromagnetic radiation
Navier-Stokes equations	describes fluid motion
Heat equation	describes heat flow
Wave equation	describes wave motion
Schrödinger equation	describes atoms, molecules and crystals
Stress-strain equations	describes elastic materials
Black-Scholes models	used for pricing financial options
Predator-prey equations	describes ecosystem populations
Einstein's equations	connects gravity and geometry
Ludwig-Jones-Holling's equation	models spruce budworm/Balsam fir ecosystem
Zeeman's model	models heart beats and nerve impulses
Sherman-Rinzel-Keizer model	for electrical activity in Pancreatic β -cells
Hodgkin-Huxley equations	models nerve action potentials

We are just going to scratch the surface of the study of differential equations. Most universities offer half a dozen different undergraduate courses on various aspects of differential equations. We will just look at one special, but important, type of equation.

Separate and integrate

Definition 2.4.1

A **separable differential equation** is an equation for a function $y(x)$ of the form

$$\frac{dy}{dx}(x) = f(x) g(y(x))$$

We'll start by developing a recipe for solving separable differential equations. Then we'll look at many examples. Usually one suppresses the argument of $y(x)$ and writes the equation ¹

$$\frac{dy}{dx} = f(x) g(y)$$

and solves such an equation by cross multiplying/dividing to get all of the y 's, including the dy on one side of the equation and all of the x 's, including the dx , on the other side of the equation.

$$\frac{dy}{g(y)} = f(x) dx$$

(We are of course assuming that $g(y)$ is nonzero.) Then you integrate both sides

$$\int \frac{dy}{g(y)} = \int f(x) dx \quad (\star)$$

This looks illegal, and indeed is illegal — $\frac{dy}{dx}$ is not a fraction. But we'll now see that the answer is still correct. This procedure is simply a mnemonic device to help you remember that answer (★).

- Our goal is to find all functions $y(x)$ that obey $\frac{dy}{dx}(x) = f(x)g(y(x))$.
- Assuming that g is nonzero,

$$\begin{aligned}
 y'(x) = f(x)g(y(x)) &\iff \frac{y'(x)}{g(y(x))} = f(x) \iff \int \frac{y'(x)}{g(y(x))} dx = \int f(x) dx \\
 &\iff \int \frac{dy}{g(y)} \Big|_{y=y(x)} = \int f(x) dx \\
 &\text{with the substitution } y = y(x), dy = y'(x) dx
 \end{aligned}$$

- That's our answer (★) again.

Let $G(y)$ be an antiderivative of $\frac{1}{g(y)}$ (i.e. $G'(y) = \frac{1}{g(y)}$) and $F(x)$ be an antiderivative of $f(x)$ (i.e. $F'(x) = f(x)$). If we reinstate the argument of y , (★) is

$$G(y(x)) = F(x) + C \tag{2.4.1}$$

Observe that the solution equation (2.4.1) contains an arbitrary constant, C . The value of this arbitrary constant *can not* be determined by the differential equation. You need additional data to determine it. Often this data consists of the value of the unknown function for one value of x . That is, often the problem you have to solve is of the form

$$\frac{dy}{dx}(x) = f(x)g(y(x)) \quad y(x_0) = y_0$$

where $f(x)$ and $g(y)$ are given functions and x_0 and y_0 are given numbers. This type of problem is called an “initial value problem”. It is solved by first using the method above to find the general solution to the differential equation, including the arbitrary constant C , and then using the “initial condition” $y(x_0) = y_0$ to determine the value of C . We'll see examples of this shortly.

✓ Example 2.4.2 A separable warm-up

The differential equation

$$\frac{dy}{dx} = xe^{-y}$$

is separable, and we now find all of its solutions by using our mnemonic device. We start by cross-multiplying so as to move all y 's to the left hand side and all x 's to the right hand side.

$$e^y dy = x dx$$

Then we integrate both sides

$$\int e^y dy = \int x dx \iff e^y = \frac{x^2}{2} + C$$

The C on the right hand side contains both the arbitrary constant for the indefinite integral $\int e^y dy$ and the arbitrary constant for the indefinite integral $\int x dx$. Finally, we solve for y , which is really a function of x .

$$y(x) = \log\left(\frac{x^2}{2} + C\right)$$

Recall that we are using \log to refer to the natural (base e) logarithm.

Note that C is an arbitrary constant. It can take any value. It cannot be determined by the differential equation itself. In applications C is usually determined by a requirement that y take some prescribed value (determined by the application) when x is some prescribed value. For example, suppose that we wish to find a function $y(x)$ that obeys both

$$\frac{dy}{dx} = xe^{-y} \quad \text{and} \quad y(0) = 1$$

We know that, to have $\frac{dy}{dx} = xe^{-y}$ satisfied, we must have $y(x) = \log\left(\frac{x^2}{2} + C\right)$, for some constant C . To also have $y(0) = 1$, we must have

$$1 = y(0) = \log\left(\frac{x^2}{2} + C\right)\Big|_{x=0} = \log C \iff \log C = 1 \iff C = e$$

So our final solution is $y(x) = \log\left(\frac{x^2}{2} + e\right)$.

✓ Example 2.4.3 A little more warm-up

Let a and b be any two constants. We'll now solve the family of differential equations

$$\frac{dy}{dx} = a(y - b)$$

using our mnemonic device.

$$\begin{aligned} \frac{dy}{y-b} = a \, dx &\implies \int \frac{dy}{y-b} = \int a \, dx \\ &\implies \log|y-b| = ax + c \implies |y-b| = e^{ax+c} = e^c e^{ax} \\ &\implies y-b = C e^{ax} \end{aligned}$$

where C is either $+e^c$ or $-e^c$. Note that as c runs over all real numbers, $+e^c$ runs over all strictly positive real numbers and $-e^c$ runs over all strictly negative real numbers. So, so far, C can be any real number except 0. But we were a bit sloppy here. We implicitly assumed that $y - b$ was nonzero, so that we could divide it across. None-the-less, the constant function $y = b$, which corresponds to $C = 0$, is a perfectly good solution — when y is the constant function $y = b$, both $\frac{dy}{dx}$ and $a(y - b)$ are zero. So the general solution to $\frac{dy}{dx} = a(y - b)$ is $y(x) = C e^{ax} + b$, where the constant C can be any real number. Note that when $y(x) = C e^{ax} + b$ we have $y(0) = C + b$. So $C = y(0) - b$ and the general solution is

$$y(x) = \{y(0) - b\} e^{ax} + b$$

This is worth stating as a theorem.

🔗 Theorem 2.4.4

Let a and b be constants. The differentiable function $y(x)$ obeys the differential equation

$$\frac{dy}{dx} = a(y - b)$$

if and only if

$$y(x) = \{y(0) - b\} e^{ax} + b$$

✓ Example 2.4.5 Solve $\frac{dy}{dx} = y^2$

Solve $\frac{dy}{dx} = y^2$

Solution: When $y \neq 0$,

$$\frac{dy}{dx} = y^2 \implies \frac{dy}{y^2} = dx \implies \frac{y^{-1}}{-1} = x + C \implies y = -\frac{1}{x + C}$$

When $y = 0$, this computation breaks down because $\frac{dy}{y^2}$ contains a division by 0. We can check if the function $y(x) = 0$ satisfies the differential equation by just subbing it in:

$$y(x) = 0 \implies y'(x) = 0, y(x)^2 = 0 \implies y'(x) = y(x)^2$$

So $y(x) = 0$ is a solution and the full solution is

$$y(x) = 0 \text{ or } y(x) = -\frac{1}{x+C}, \text{ for any constant } C$$

✓ Example 2.4.6 A falling raindrop

When a raindrop falls it increases in size so that its mass $m(t)$, is a function of time t . The rate of growth of mass, i.e. $\frac{dm}{dt}$, is $km(t)$ for some positive constant k . According to Newton's law of motion, $\frac{d}{dt}(mv) = gm$, where v is the velocity of the raindrop (with v being positive for downward motion) and g is the acceleration due to gravity. Find the terminal velocity, $\lim_{t \rightarrow \infty} v(t)$, of a raindrop.

Solution: In this problem we have two unknown functions, $m(t)$ and $v(t)$, and two differential equations, $\frac{dm}{dt} = km$ and $\frac{d}{dt}(mv) = gm$. The first differential equation, $\frac{dm}{dt} = km$, involves only $m(t)$, not $v(t)$, so we use it to determine $m(t)$. By Theorem 2.4.4, with $b = 0$, $a = k$, y replaced by m and x replaced by t ,

$$\frac{dm}{dt} = km \implies m(t) = m(0)e^{kt}$$

Now that we know $m(t)$ (except for the value of the constant $m(0)$), we can substitute it into the second differential equation, which we can then use to determine the remaining unknown function $v(t)$. Observe that the second equation, $\frac{d}{dt}(mv) = gm(t) = gm(0)e^{kt}$ tells that the derivative of the function $y(t) = m(t)v(t)$ is $gm(0)e^{kt}$. So $y(t)$ is just an antiderivative of $gm(0)e^{kt}$.

$$\frac{dy}{dt} = gm(t) = gm(0)e^{kt} \implies y(t) = \int gm(0)e^{kt} dt = gm(0)\frac{e^{kt}}{k} + C$$

Now that we know $y(t) = m(t)v(t) = m(0)e^{kt}v(t)$, we can get $v(t)$ just by dividing out the $m(0)e^{kt}$.

$$\begin{aligned} y(t) = gm(0)\frac{e^{kt}}{k} + C &\implies m(0)e^{kt}v(t) = gm(0)\frac{e^{kt}}{k} + C \\ &\implies v(t) = \frac{g}{k} + \frac{C}{m(0)e^{kt}} \end{aligned}$$

Our solution, $v(t)$, contains two arbitrary constants, namely C and $m(0)$. They will be determined by, for example, the mass and velocity at time $t = 0$. But since we are only interested in the terminal velocity $\lim_{t \rightarrow \infty} v(t)$, we don't need to know C and $m(0)$. Since $k > 0$, $\lim_{t \rightarrow \infty} \frac{C}{e^{kt}} = 0$ and the terminal velocity $\lim_{t \rightarrow \infty} v(t) = \frac{g}{k}$.

✓ Example 2.4.7 Intravenous glucose

A glucose solution is administered intravenously into the bloodstream at a constant rate r . As the glucose is added, it is converted into other substances at a rate that is proportional to the concentration at that time. The concentration, $C(t)$, of the glucose in the bloodstream at time t obeys the differential equation

$$\frac{dC}{dt} = r - kC$$

where k is a positive constant of proportionality.

1. Express $C(t)$ in terms of k and $C(0)$.
2. Find $\lim_{t \rightarrow \infty} C(t)$.

Solution: (a) Since $r - kC = -k\left(C - \frac{r}{k}\right)$ the given equation is

$$\frac{dC}{dt} = -k\left(C - \frac{r}{k}\right)$$

which is of the form solved in Theorem 2.4.4 with $a = -k$ and $b = \frac{r}{k}$. So the solution is

$$C(t) = \frac{r}{k} + \left(C(0) - \frac{r}{k}\right)e^{-kt}$$

For any $k > 0$, $\lim_{t \rightarrow \infty} e^{-kt} = 0$. Consequently, for any $C(0)$ and any $k > 0$, $\lim_{t \rightarrow \infty} C(t) = \frac{r}{k}$. We could have predicted this limit without solving for $C(t)$. If we assume that $C(t)$ approaches some equilibrium value C_e as t approaches infinity, then taking the limits of both sides of $\frac{dC}{dt} = r - kC$ as $t \rightarrow \infty$ gives

$$0 = r - kC_e \implies C_e = \frac{r}{k}$$

Optional — Carbon Dating

Scientists can determine the age of objects containing organic material by a method called *carbon dating* or *radiocarbon dating*². The bombardment of the upper atmosphere by cosmic rays converts nitrogen to a radioactive isotope of carbon, ^{14}C , with a half-life of about 5730 years. Vegetation absorbs carbon dioxide from the atmosphere through photosynthesis and animals acquire ^{14}C by eating plants. When a plant or animal dies, it stops replacing its carbon and the amount of ^{14}C begins to decrease through radioactive decay. Therefore the level of radioactivity also decreases. More precisely, let $Q(t)$ denote the amount of ^{14}C in the plant or animal t years after it dies. The number of radioactive decays per unit time, at time t , is proportional to the amount of ^{14}C present at time t , which is $Q(t)$. Thus

Equation 2.4.8 Radioactive decay

$$\frac{dQ}{dt}(t) = -kQ(t)$$

Here k is a constant of proportionality that is determined by the half-life. We shall explain what half-life is, and also determine the value of k , in Example 2.4.9, below.

Before we do so, let's think about the sign in equation (2.4.8).

- Recall that $Q(t)$ denotes a quantity, namely the amount of ^{14}C present at time t . There cannot be a negative amount of ^{14}C . Nor can this quantity be zero. (We would not use carbon dating when there is no ^{14}C present.) Consequently, $Q(t) > 0$.
- As the time t increases, $Q(t)$ decreases, because ^{14}C is being continuously converted into ^{14}N by radioactive decay³. Thus $\frac{dQ}{dt}(t) < 0$.
- The signs $Q(t) > 0$ and $\frac{dQ}{dt}(t) < 0$ are consistent with (2.4.8) provided the constant of proportionality $k > 0$.
- In (2.4.8), we chose to call the constant of proportionality “ $-k$ ”. We did so in order to make $k > 0$. We could just as well have chosen to call the constant of proportionality “ K ”. That is, we could have replaced equation (2.4.8) by $\frac{dQ}{dt}(t) = KQ(t)$. The constant of proportionality K would have to be negative, (and K and k would be related by $K = -k$).

Example 2.4.9 Half life and the constant k

In this example, we determine the value of the constant of proportionality k in (2.4.8) that corresponds to the half-life of ^{14}C , which is 5730 years.

- Imagine that some plant or animal contains a quantity Q_0 of ^{14}C at its time of death. Let's choose the zero point of time $t = 0$ to be the instant that the plant or animal died.
- Denote by $Q(t)$ the amount of ^{14}C in the plant or animal t years after it died. Then $Q(t)$ must obey both (2.4.8) and $Q(0) = Q_0$.
- Theorem 2.4.4, with $b = 0$ and $a = -k$, then tells us that $Q(t) = Q_0 e^{-kt}$ for all $t \geq 0$.

- By definition, the half-life of ^{14}C is the length of time that it takes for half of the ^{14}C to decay. That is, the half-life $t_{1/2}$ is determined by

$$Q(t_{1/2}) = \frac{1}{2}Q(0) = \frac{1}{2}Q_0 \quad \text{but we know that } Q(t) = Q_0e^{-kt}$$

$$Q_0e^{-kt_{1/2}} = \frac{1}{2}Q_0 \quad \text{now cancel } Q_0$$

$$e^{-kt_{1/2}} = \frac{1}{2}$$

Taking the logarithm of both sides gives

$$-kt_{1/2} = \log \frac{1}{2} = -\log 2 \implies k = \frac{\log 2}{t_{1/2}}$$

Recall that, in this text, we use $\log x$ to indicate the natural logarithm. That is,

$$\log x = \log_e x = \ln x$$

We are told that, for ^{14}C , the half-life $t_{1/2} = 5730$, so

$$k = \frac{\log 2}{5730} = 0.000121 \quad \text{to 6 decimal places}$$

From the work in the above example we have accumulated enough new facts to make a corollary to Theorem 2.4.4.

Corollary 2.4.10

The function $Q(t)$ satisfies the equation

$$\frac{dQ}{dt} = -kQ(t)$$

if and only if

$$Q(t) = Q(0)e^{-kt}$$

The half-life is defined to be the time $t_{1/2}$ which obeys

$$Q(t_{1/2}) = \frac{1}{2}Q(0)$$

The half-life is related to the constant k by

$$t_{1/2} = \frac{\log 2}{k}$$

Now here is a typical problem that is solved using Corollary 2.4.10.

Example 2.4.11 The age of a piece of parchment

A particular piece of parchment contains about 64% as much ^{14}C as plants do today. Estimate the age of the parchment.

Solution: Let $Q(t)$ denote the amount of ^{14}C in the parchment t years after it was first created. By equation (2.4.8) and Example 2.4.9

$$\frac{dQ}{dt}(t) = -kQ(t) \quad \text{with } k = \frac{\log 2}{5730} = 0.000121$$

By Corollary 2.4.10

$$Q(t) = Q(0)e^{-kt}$$

The time at which $Q(t)$ reaches 0.64 $Q(0)$ is determined by

$$\begin{aligned}
 Q(t) &= 0.64 Q(0) && \text{but } Q(t) = Q(0) e^{-kt} \\
 Q(0) e^{-kt} &= 0.64 Q(0) && \text{cancel } Q(0) \\
 e^{-kt} &= 0.64 && \text{take logarithms} \\
 -kt &= \log 0.64 \\
 t &= \frac{\log 0.64}{-k} = \frac{\log 0.64}{-0.000121} = 3700 && \text{to 2 significant digits}
 \end{aligned}$$

That is, the parchment ⁴ is about 37 centuries old.

We have stated that the half-life of ¹⁴C is 5730 years. How can this be determined? We can explain this using the following example.

✓ Example 2.4.12 Half life of implausium

A scientist in a B-grade science fiction film is studying a sample of the rare and fictitious element, implausium. With great effort he has produced a sample of pure implausium. The next day — 17 hours later — he comes back to his lab and discovers that his sample is now only 37% pure. What is the half-life of the element?

Solution: We can again set up our problem using Corollary 2.4.10. Let $Q(t)$ denote the quantity of implausium at time t , measured in hours. Then we know

$$Q(t) = Q(0) \cdot e^{-kt}$$

We also know that

$$Q(17) = 0.37Q(0).$$

That enables us to determine k via

$$\begin{aligned}
 Q(17) = 0.37Q(0) &= Q(0)e^{-17k} && \text{divide both sides by } Q(0) \\
 0.37 &= e^{-17k}
 \end{aligned}$$

and so

$$k = -\frac{\log 0.37}{17} = 0.05849$$

We can then convert this to the half life using Corollary 2.4.10:

$$t_{1/2} = \frac{\log 2}{k} \approx 11.85 \text{ hours}$$

While this example is entirely fictitious, one really can use this approach to measure the half-life of materials.

Optional — Newton's Law of Cooling

Newton's law of cooling says:

- The rate of change of temperature of an object is proportional to the difference in temperature between the object and its surroundings. The temperature of the surroundings is sometimes called the ambient temperature.

If we denote by $T(t)$ the temperature of the object at time t and by A the temperature of its surroundings, Newton's law of cooling says that there is some constant of proportionality, K , such that

🔗 Equation 2.4.13 Newton's law of cooling

$$\frac{dT}{dt}(t) = K[T(t) - A]$$

This mathematical model of temperature change works well when studying a small object in a large, fixed temperature, environment. For example, a hot cup of coffee in a large room ⁵. Let's start by thinking a little about the sign of the constant of

proportionality. At any time t , there are three possibilities.

- If $T(t) > A$, that is, if the body is warmer than its surroundings, we would expect heat to flow from the body into its surroundings and so we would expect the body to cool off so that $\frac{dT}{dt}(t) < 0$. For this expectation to be consistent with (2.4.13), we need $K < 0$.
- If $T(t) < A$, that is the body is cooler than its surroundings, we would expect heat to flow from the surroundings into the body and so we would expect the body to warm up so that $\frac{dT}{dt}(t) > 0$. For this expectation to be consistent with (2.4.13), we again need $K < 0$.
- Finally if $T(t) = A$, that is the body and its environment have the same temperature, we would not expect any heat to flow between the two and so we would expect that $\frac{dT}{dt}(t) = 0$. This does not impose any condition on K .

In conclusion, we would expect $K < 0$. Of course, we could have chosen to call the constant of proportionality $-k$, rather than K . Then the differential equation would be $\frac{dT}{dt} = -k(T - A)$ and we would expect $k > 0$.

✓ Example 2.4.14 Warming iced tea

The temperature of a glass of iced tea is initially 5° . After 5 minutes, the tea has heated to 10° in a room where the air temperature is 30° .

1. Determine the temperature as a function of time.
2. What is the temperature after 10 minutes?
3. Determine when the tea will reach a temperature of 20° .

Solution: (a)

- Denote by $T(t)$ the temperature of the tea t minutes after it was removed from the fridge, and let $A = 30$ be the ambient temperature.
- By Newton's law of cooling,

$$\frac{dT}{dt} = K(T - A) = K(T - 30)$$

for some, as yet unknown, constant of proportionality K .

- By Theorem 2.4.4 with $a = K$ and $b = 30$,

$$T(t) = [T(0) - 30] e^{Kt} + 30 = 30 - 25e^{Kt}$$

since the initial temperature $T(0) = 5$.

- This solution is not complete because it still contains an unknown constant, namely K . We have not yet used the given data that $T(5) = 10$. We can use it to determine K . At $t = 5$,

$$\begin{aligned} T(5) = 30 - 25e^{5K} = 10 &\implies e^{5K} = \frac{20}{25} \implies 5K = \log \frac{20}{25} \\ &\implies K = \frac{1}{5} \log \frac{4}{5} = -0.044629 \end{aligned}$$

to six decimal places.

- (b) To find the temperature at 10 minutes we can just use the solution we have determined above.

$$\begin{aligned} T(10) &= 30 - 25e^{10K} \\ &= 30 - 25e^{10 \times \frac{1}{5} \log \frac{4}{5}} \\ &= 30 - 25e^{2 \log \frac{4}{5}} = 30 - 25e^{\log \frac{16}{25}} \\ &= 30 - 16 = 14^\circ \end{aligned}$$

- (c) The temperature is 20° when

$$30 - 25e^{Kt} = 20 \implies e^{Kt} = \frac{10}{25} \implies Kt = \log \frac{10}{25}$$

$$\implies t = \frac{1}{K} \log \frac{2}{5} = 20.5 \text{ min}$$

to one decimal place.

✓ Example 2.4.15 Temperature back in time

A dead body is discovered at 3:45pm in a room where the temperature is 20°C. At that time the temperature of the body is 27°C. Two hours later, at 5:45pm, the temperature of the body is 25.3°C. What was the time of death? Note that the normal (adult human) body temperature is 37°C.

Solution: We will assume that the body's temperature obeys Newton's law of cooling.

- Denote by $T(t)$ the temperature of the body at time t , with $t = 0$ corresponding to 3:45pm. We wish to find the time of death — call it t_d .
- There is a lot of data in the statement of the problem. We are told
 1. the ambient temperature: $A = 20$
 2. the temperature of the body when discovered: $T(0) = 27$
 3. the temperature of the body 2 hours later: $T(2) = 25.3$
 4. assuming the person was a healthy adult right up until he died, the temperature at the time of death: $T(t_d) = 37$.
- Theorem 2.4.4 with $a = K$ and $b = A = 20$

$$T(t) = [T(0) - A] e^{Kt} + A = 20 + 7e^{Kt}$$

Two unknowns remain, K and t_d .

- We can find the first, K , by using the condition (3), which says $T(2) = 25.3$.

$$25.3 = T(2) = 20 + 7e^{2K} \implies 7e^{2K} = 5.3 \implies 2K = \log \left(\frac{5.3}{7} \right)$$

$$\implies K = \frac{1}{2} \log \left(\frac{5.3}{7} \right) = -0.139$$

- Finally, t_d is determined by the condition (4).

$$37 = T(t_d) = 20 + 7e^{-0.139t_d} \implies e^{-0.139t_d} = \frac{17}{7}$$

$$\implies -0.139t_d = \log \left(\frac{17}{7} \right)$$

$$\implies t_d = -\frac{1}{0.139} \log \left(\frac{17}{7} \right) = -6.38$$

to two decimal places. Now 6.38 hours is 6 hours and $0.38 \times 60 = 23$ minutes. So the time of death was 6 hours and 23 minutes before 3:45pm, which is 9:22am.

A slightly tricky example — we need to determine the ambient temperature from three measurements at different times.

✓ Example 2.4.16 Finding the ambient temperature

A glass of room-temperature water is carried out onto a balcony from an apartment where the temperature is 22°C. After one minute the water has temperature 26°C and after two minutes it has temperature 28°C. What is the outdoor temperature?

Solution: We will assume that the temperature of the thermometer obeys Newton's law of cooling.

- Let A be the outdoor temperature and $T(t)$ be the temperature of the water t minutes after it is taken outside.
- By Newton's law of cooling,

$$T(t) = A + (T(0) - A)e^{Kt}$$

Theorem 2.4.4 with $a = K$ and $b = A$. Notice there are 3 unknowns here — A , $T(0)$ and K — so we need three pieces of information to find them all.

- We are told $T(0) = 22$, so

$$T(t) = A + (22 - A)e^{Kt}.$$

- We are also told $T(1) = 26$, which gives

$$26 = A + (22 - A)e^K \quad \text{rearrange things}$$

$$e^K = \frac{26 - A}{22 - A}$$

- Finally, $T(2) = 28$, so

$$28 = A + (22 - A)e^{2K}$$

$$e^{2K} = \frac{28 - A}{22 - A} \quad \text{rearrange}$$

but $e^K = \frac{26 - A}{22 - A}$, so

$$\left(\frac{26 - A}{22 - A}\right)^2 = \frac{28 - A}{22 - A} \quad \text{multiply through by } (22 - A)^2$$

$$(26 - A)^2 = (28 - A)(22 - A)$$

We can expand out both sides and collect up terms to get

$$\underbrace{26^2}_{=676} - 52A + A^2 = \underbrace{28 \times 22}_{=616} - 50A + A^2$$

$$60 = 2A$$

$$30 = A$$

So the temperature outside is 30° .

Optional — Population Growth

Suppose that we wish to predict the size $P(t)$ of a population as a function of the time t . In the most naive model of population growth, each couple produces β offspring (for some constant β) and then dies. Thus over the course of one generation $\beta \frac{P(t)}{2}$ children are produced and $P(t)$ parents die so that the size of the population grows from $P(t)$ to

$$P(t + t_g) = \underbrace{P(t) + \beta \frac{P(t)}{2}}_{\text{parents+offspring}} - \underbrace{P(t)}_{\text{parents die}} = \frac{\beta}{2} P(t)$$

where t_g denotes the lifespan of one generation. The rate of change of the size of the population per unit time is

$$\frac{P(t + t_g) - P(t)}{t_g} = \frac{1}{t_g} \left[\frac{\beta}{2} P(t) - P(t) \right] = bP(t)$$

where $b = \frac{\beta - 2}{2t_g}$ is the net birthrate per member of the population per unit time. If we approximate

$$\frac{P(t + t_g) - P(t)}{t_g} \approx \frac{dP}{dt}(t)$$

we get the differential equation

Equation 2.4.17 Population growth

$$\frac{dP}{dt} = bP(t)$$

By Corollary 2.4.10, with $-k$ replaced by b ,

Equation 2.4.18 Malthusian growth model

$$P(t) = P(0) \cdot e^{bt}$$

This is called the Malthusian ⁶ growth model. It is, of course, very simplistic. One of its main characteristics is that, since $P(t+T) = P(0) \cdot e^{b(t+T)} = P(t) \cdot e^{bT}$, every time you *add* T to the time, the population size is *multiplied* by e^{bT} . In particular, the population size doubles every $\frac{\log 2}{b}$ units of time. The Malthusian growth model can be a reasonably good model only when the population size is very small compared to its environment ⁷. A more sophisticated model of population growth, that takes into account the “carrying capacity of the environment” is considered below.

✓ Example 2.4.19 A rough estimate of the earth's population

In 1927 the population of the world was about 2 billion. In 1974 it was about 4 billion. Estimate when it reached 6 billion. What will the population of the world be in 2100, assuming the Malthusian growth model?

Solution: We follow our usual pattern for dealing with such problems.

- Let $P(t)$ be the world's population, in billions, t years after 1927. Note that 1974 corresponds to $t = 1974 - 1927 = 47$.
- We are assuming that $P(t)$ obeys equation (2.4.17). So, by (2.4.18)

$$P(t) = P(0) \cdot e^{bt}$$

Notice that there are 2 unknowns here — b and $P(0)$ — so we need two pieces of information to find them.

- We are told $P(0) = 2$, so

$$P(t) = 2 \cdot e^{bt}$$

- We are also told $P(47) = 4$, which gives

$$\begin{aligned} 4 &= 2 \cdot e^{47b} && \text{clean up} \\ e^{47b} &= 2 && \text{take the log and clean up} \\ b &= \frac{\log 2}{47} = 0.0147 && \text{to 3 decimal places} \end{aligned}$$

- We now know $P(t)$ completely, so we can easily determine the predicted population ⁸ in 2100, i.e. at $t = 2100 - 1927 = 173$.

$$P(173) = 2e^{173b} = 2e^{173 \times 0.0147} = 12.7 \text{ billion}$$

- Finally, our crude model predicts that the population is 6 billion at the time t that obeys

$$\begin{aligned} P(t) &= 2e^{bt} = 6 && \text{clean up} \\ e^{bt} &= 3 && \text{take the log and clean up} \\ t &= \frac{\log 3}{b} = 47 \frac{\log 3}{\log 2} = 74.5 \end{aligned}$$

which corresponds ⁹ to the middle of 2001.

Logistic growth adds one more wrinkle to the simple population model. It assumes that the population only has access to limited resources. As the size of the population grows the amount of food available to each member decreases. This in turn causes the net birth rate b to decrease. In the logistic growth model $b = b_0 \left(1 - \frac{P}{K}\right)$, where K is called the carrying capacity of the environment, so that

$$P'(t) = b_0 \left(1 - \frac{P(t)}{K}\right) P(t)$$

This is a separable differential equation and we can solve it explicitly. We shall do so shortly. See Example 2.4.20, below. But, before doing that, we'll see what we can learn about the behaviour of solutions to differential equations like this without finding

formulae for the solutions. It turns out that we can learn a lot just by watching the sign of $P'(t)$. For concreteness, we'll look at solutions of the differential equation

$$\frac{dP}{dt}(t) = (6000 - 3P(t)) P(t)$$

We'll sketch the graphs of four functions $P(t)$ that obey this equation.

- For the first function, $P(0) = 0$.
- For the second function, $P(0) = 1000$.
- For the third function, $P(0) = 2000$.
- For the fourth function, $P(0) = 3000$.

The sketches will be based on the observation that $(6000 - 3P) P = 3(2000 - P) P$

- is zero for $P = 0, 2000$,
- is strictly positive for $0 < P < 2000$ and
- is strictly negative for $P > 2000$.

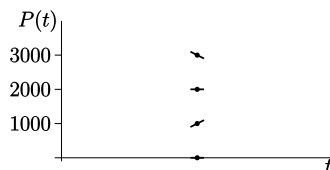
Consequently

$$\frac{dP}{dt}(t) \begin{cases} = 0 & \text{if } P(t) = 0 \\ > 0 & \text{if } 0 < P(t) < 2000 \\ = 0 & \text{if } P(t) = 2000 \\ < 0 & \text{if } P(t) > 2000 \end{cases}$$

Thus if $P(t)$ is some function that obeys $\frac{dP}{dt}(t) = (6000 - 3P(t))P(t)$, then as the graph of $P(t)$ passes through the point $(t, P(t))$

$$\text{the graph has } \begin{cases} \text{slope zero,} & \text{i.e. is horizontal, } \backslash \backslash \text{ if } P(t) = 0 \\ \text{positive slope,} & \text{i.e. is increasing, } \backslash \backslash \text{ if } 0 < P(t) < 2000 \\ \text{slope zero,} & \text{i.e. is horizontal, } \backslash \backslash \text{ if } P(t) = 2000 \\ \text{negative slope,} & \text{i.e. is decreasing, } \backslash \backslash \text{ if } P(t) > 2000 \end{cases}$$

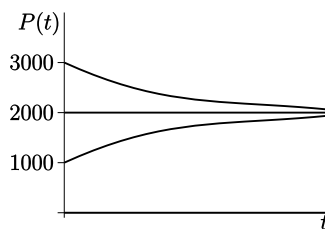
as illustrated in the figure



As a result,

- if $P(0) = 0$, the graph starts out horizontally. In other words, as t starts to increase, $P(t)$ remains at zero, so the slope of the graph remains at zero. The population size remains zero for all time. As a check, observe that the function $P(t) = 0$ obeys $\frac{dP}{dt}(t) = (6000 - 3P(t))P(t)$ for all t .
- Similarly, if $P(0) = 2000$, the graph again starts out horizontally. So $P(t)$ remains at 2000 and the slope remains at zero. The population size remains 2000 for all time. Again, the function $P(t) = 2000$ obeys $\frac{dP}{dt}(t) = (6000 - 3P(t))P(t)$ for all t .
- If $P(0) = 1000$, the graph starts out with positive slope. So $P(t)$ increases with t . As $P(t)$ increases towards 2000, the slope $(6000 - 3P(t))P(t)$, while remaining positive, gets closer and closer to zero. As the graph approaches height 2000, it becomes more and more horizontal. The graph cannot actually cross from below 2000 to above 2000, because to do so it would have to have strictly positive slope for some value of P above 2000, which is not allowed.
- If $P(0) = 3000$, the graph starts out with negative slope. So $P(t)$ decreases with t . As $P(t)$ decreases towards 2000, the slope $(6000 - 3P(t))P(t)$, while remaining negative, gets closer and closer to zero. As the graph approaches height 2000, it becomes more and more horizontal. The graph cannot actually cross from above 2000 to below 2000, because to do so it would have to have negative slope for some value of P below 2000, which is not allowed.

These curves are sketched in the figure below. We conclude that for any initial population size $P(0)$, except $P(0) = 0$, the population size approaches 2000 as $t \rightarrow \infty$.



Now we'll do an example in which we explicitly solve the logistic growth equation.

✓ Example 2.4.20 Population predictions using logistic growth

In 1986, the population of the world was 5 billion and was increasing at a rate of 2% per year. Using the logistic growth model with an assumed maximum population of 100 billion, predict the population of the world in the years 2000, 2100 and 2500.

Solution: Let $y(t)$ be the population of the world, in billions of people, at time $1986 + t$. The logistic growth model assumes

$$y' = ay(K - y)$$

where K is the carrying capacity and $a = \frac{b_0}{K}$.

First we'll determine the values of the constants a and K from the given data.

- We know that, if at time zero the population is below K , then as time increases the population increases, approaching the limit K as t tends to infinity. So in this problem K is the maximum population. That is, $K = 100$.
- We are also told that, at time zero, the percentage rate of change of population, $100 \frac{y'}{y}$, is 2, so that, at time zero, $\frac{y'}{y} = 0.02$. But, from the differential equation, $\frac{y'}{y} = a(K - y)$. Hence at time zero, $0.02 = a(100 - 5)$, so that $a = \frac{2}{9500}$.

We now know a and K and can solve the (separable) differential equation

$$\begin{aligned} \frac{dy}{dt} = ay(K - y) &\implies \frac{dy}{y(K - y)} = a dt \\ &\implies \int \frac{1}{K} \left[\frac{1}{y} - \frac{1}{y - K} \right] dy = \int a dt \\ &\implies \frac{1}{K} [\log |y| - \log |y - K|] = at + C \\ &\implies \log \frac{|y|}{|y - K|} = aKt + CK \implies \left| \frac{y}{y - K} \right| = De^{aKt} \end{aligned}$$

with $D = e^{CK}$. We know that y remains between 0 and K , so that $\left| \frac{y}{y - K} \right| = \frac{y}{K - y}$ and our solution obeys

$$\frac{y}{K - y} = De^{aKt}$$

At this stage, we know the values of the constants a and K , but not the value of the constant D . We are given that at $t = 0$, $y = 5$. Subbing in this, and the values of K and a ,

$$\frac{5}{100 - 5} = De^0 \implies D = \frac{5}{95}$$

So the solution obeys the algebraic equation

$$\frac{y}{100 - y} = \frac{5}{95} e^{2t/95}$$

which we can solve to get y as a function of t .

$$\begin{aligned}
 y &= (100 - y) \frac{5}{95} e^{2t/95} \implies 95y = (500 - 5y) e^{2t/95} \\
 &\implies (95 + 5e^{2t/95})y = 500e^{2t/95} \\
 &\implies y = \frac{500e^{2t/95}}{95 + 5e^{2t/95}} = \frac{100e^{2t/95}}{19 + e^{2t/95}} = \frac{100}{1 + 19e^{-2t/95}}
 \end{aligned}$$

Finally,

- In the year 2000, $t = 14$ and $y = \frac{100}{1 + 19e^{-28/95}} \approx 6.6$ billion.
- In the year 2100, $t = 114$ and $y = \frac{100}{1 + 19e^{-228/95}} \approx 36.7$ billion.
- In the year 2200, $t = 514$ and $y = \frac{100}{1 + 19e^{-1028/95}} \approx 100$ billion.

Optional — Mixing Problems

✓ Example 2.4.21 Dissolving salt

At time $t = 0$, where t is measured in minutes, a tank with a 5-litre capacity contains 3 litres of water in which 1 kg of salt is dissolved. Fresh water enters the tank at a rate of 2 litres per minute and the fully mixed solution leaks out of the tank at the varying rate of $2t$ litres per minute.

1. Determine the volume of solution $V(t)$ in the tank at time t .
2. Determine the amount of salt $Q(t)$ in solution when the amount of water in the tank is at maximum.

Solution¹⁰: (a) The rate of change of the volume in the tank, at time t , is $2 - 2t$, because water is entering at a rate 2 and solution is leaking out at a rate $2t$. Thus

$$\frac{dV}{dt} = 2 - 2t \implies dV = (2 - 2t) dt \implies V = \int (2 - 2t) dt = 2t - t^2 + C$$

at least until $V(t)$ reaches either the capacity of the tank or zero. When $t = 0$, $V = 3$ so $C = 3$ and $V(t) = 3 + 2t - t^2$. Observe that $V(t)$ is at a maximum when $\frac{dV}{dt} = 2 - 2t = 0$, or $t = 1$.

(b) In the very short time interval from time t to time $t + dt$, $2t dt$ litres of brine leaves the tank. That is, the fraction $\frac{2t dt}{V(t)}$ of the total salt in the tank, namely $Q(t) \frac{2t dt}{V(t)}$ kilograms, leaves. Thus salt is leaving the tank at the rate

$$\frac{Q(t) \frac{2t dt}{V(t)}}{dt} = \frac{2tQ(t)}{V(t)} = \frac{2tQ(t)}{3 + 2t - t^2} \text{ kilograms per minute}$$

so

$$\begin{aligned}
 \frac{dQ}{dt} &= -\frac{2tQ(t)}{3 + 2t - t^2} \implies \frac{dQ}{Q} = -\frac{2t}{3 + 2t - t^2} dt = -\frac{2t}{(3 - t)(1 + t)} dt \\
 &= \left[\frac{3/2}{t - 3} + \frac{1/2}{t + 1} \right] dt \\
 &\implies \log Q = \frac{3}{2} \log |t - 3| + \frac{1}{2} \log |t + 1| + C
 \end{aligned}$$

We are interested in the time interval $0 \leq t \leq 1$. In this time interval $|t - 3| = 3 - t$ and $|t + 1| = t + 1$ so

$$\log Q = \frac{3}{2} \log(3 - t) + \frac{1}{2} \log(t + 1) + C$$

At $t = 0$, Q is 1 so

$$\begin{aligned}
 \log 1 &= \frac{3}{2} \log(3 - 0) + \frac{1}{2} \log(0 + 1) + C \\
 &\implies C = \log 1 - \frac{3}{2} \log 3 - \frac{1}{2} \log 1 = -\frac{3}{2} \log 3
 \end{aligned}$$

At $t = 1$

$$\begin{aligned}\log Q &= \frac{3}{2} \log(3-1) + \frac{1}{2} \log(1+1) - \frac{3}{2} \log 3 \\ &= 2 \log 2 - \frac{3}{2} \log 3 = \log 4 - \log 3^{\frac{3}{2}}\end{aligned}$$

so $Q = \frac{4}{3^{\frac{3}{2}}}$.

✓ Example 2.4.22 Mixing brines

A tank contains 1500 liters of brine with a concentration of 0.3 kg of salt per liter. Another brine solution, this with a concentration of 0.1 kg of salt per liter is poured into the tank at a rate of 20 li/min. At the same time, 20 li/min of the solution in the tank, which is stirred continuously, is drained from the tank.

1. How many kilograms of salt will remain in the tank after half an hour?
2. How long will it take to reduce the concentration to 0.2 kg/li?

Solution: Denote by $Q(t)$ the amount of salt in the tank at time t . In a very short time interval dt , the incoming solution adds $20 dt$ liters of a solution carrying 0.1 kg/li. So the incoming solution adds $0.1 \times 20 dt = 2 dt$ kg of salt. In the same time interval $20 dt$ liters is drained from the tank. The concentration of the drained brine is $\frac{Q(t)}{1500}$. So $\frac{Q(t)}{1500} 20 dt$ kg were removed. All together, the change in the salt content of the tank during the short time interval is

$$dQ = 2 dt - \frac{Q(t)}{1500} 20 dt = \left(2 - \frac{Q(t)}{75}\right) dt$$

The rate of change of salt content per unit time is

$$\frac{dQ}{dt} = 2 - \frac{Q(t)}{75} = -\frac{1}{75} (Q(t) - 150)$$

The solution of this equation is

$$Q(t) = \{Q(0) - 150\} e^{-t/75} + 150$$

by Theorem 2.4.4, with $a = -\frac{1}{75}$ and $b = 150$. At time 0, $Q(0) = 1500 \times 0.3 = 450$. So

$$Q(t) = 150 + 300e^{-t/75}$$

(a) At $t = 30$

$$Q(30) = 150 + 300e^{-30/75} = 351.1 \text{ kg}$$

(b) $Q(t) = 0.2 \times 1500 = 300$ kg is achieved when

$$\begin{aligned}150 + 300e^{-t/75} = 300 &\implies 300e^{-t/75} = 150 \implies e^{-t/75} = 0.5 \\ \implies -\frac{t}{75} = \log(0.5) &\implies t = -75 \log(0.5) = 51.99 \text{ min}\end{aligned}$$

Optional — Interest on Investments

Suppose that you deposit $\$P$ in a bank account at time $t = 0$. The account pays $r\%$ interest per year compounded n times per year.

- The first interest payment is made at time $t = \frac{1}{n}$. Because the balance in the account during the time interval $0 < t < \frac{1}{n}$ is $\$P$ and interest is being paid for $(\frac{1}{n})^{\text{th}}$ of a year, that first interest payment is $\frac{1}{n} \times \frac{r}{100} \times P$. After the first interest payment, the balance in the account is $P + \frac{1}{n} \times \frac{r}{100} \times P = (1 + \frac{r}{100n})P$.
- The second interest payment is made at time $t = \frac{2}{n}$. Because the balance in the account during the time interval $\frac{1}{n} < t < \frac{2}{n}$ is $(1 + \frac{r}{100n})P$ and interest is being paid for $(\frac{1}{n})^{\text{th}}$ of a year, the second interest payment is $\frac{1}{n} \times \frac{r}{100} \times (1 + \frac{r}{100n})P$. After the second interest payment, the balance in the account is $(1 + \frac{r}{100n})P + \frac{1}{n} \times \frac{r}{100} \times (1 + \frac{r}{100n})P = (1 + \frac{r}{100n})^2 P$.
- And so on.

In general, at time $t = \frac{m}{n}$ (just after the m^{th} interest payment), the balance in the account is

Equation 2.4.23 Discrete compounding interest

$$B(t) = \left(1 + \frac{r}{100n}\right)^m P = \left(1 + \frac{r}{100n}\right)^{nt} P$$

Three common values of n are 1 (interest is paid once a year), 12 (i.e. interest is paid once a month) and 365 (i.e. interest is paid daily). The limit $n \rightarrow \infty$ is called continuous compounding¹¹. Under continuous compounding, the balance at time t is

$$B(t) = \lim_{n \rightarrow \infty} \left(1 + \frac{r}{100n}\right)^{nt} P$$

You may have already seen the limit

Equation 2.4.24 A useful limit

$$\lim_{x \rightarrow 0} (1+x)^{a/x} = e^a$$

If so, you can evaluate $B(t)$ by applying 2.4.24 with $x = \frac{r}{100n}$ and $a = \frac{rt}{100}$ (so that $\frac{a}{x} = nt$). As $n \rightarrow \infty$, $x \rightarrow 0$ so that

Equation 2.4.25 Continuous compound interest

$$B(t) = \lim_{n \rightarrow \infty} \left(1 + \frac{r}{100n}\right)^{nt} P = \lim_{x \rightarrow 0} (1+x)^{a/x} P = e^a P = e^{rt/100} P$$

If you haven't seen (2.4.24) before, that's OK. In the following example, we rederive (2.4.25) using a differential equation instead of (2.4.24).

Example 2.4.26 Computing my future bank balance

Suppose, again, that you deposit $\$P$ in a bank account at time $t = 0$, and that the account pays $r\%$ interest per year compounded n times per year, and denote by $B(t)$ the balance at time t . Suppose that you have just received an interest payment at time t . Then the next interest payment will be made at time $t + \frac{1}{n}$ and will be $\frac{1}{n} \times \frac{r}{100} \times B(t) = \frac{r}{100n} B(t)$. So, calling $\frac{1}{n} = h$,

$$B(t+h) = B(t) + \frac{r}{100} B(t)h \quad \text{or} \quad \frac{B(t+h) - B(t)}{h} = \frac{r}{100} B(t)$$

To get continuous compounding we take the limit $n \rightarrow \infty$ or, equivalently, $h \rightarrow 0$. This gives

$$\lim_{h \rightarrow 0} \frac{B(t+h) - B(t)}{h} = \frac{r}{100} B(t) \quad \text{or} \quad \frac{dB}{dt}(t) = \frac{r}{100} B(t)$$

By Theorem 2.4.4, with $a = \frac{r}{100}$ and $b = 0$, (or Corollary 2.4.10 with $k = -\frac{r}{100}$),

$$B(t) = e^{rt/100} B(0) = e^{rt/100} P$$

once again.

Example 2.4.27 Double your money

1. A bank advertises that it compounds interest continuously and that it will double your money in ten years. What is the annual interest rate?
2. A bank advertises that it compounds monthly and that it will double your money in ten years. What is the annual interest rate?

Solution: (a) Let the interest rate be $r\%$ per year. If you start with $\$P$, then after t years, you have $Pe^{rt/100}$, under continuous compounding. This was equation (2.4.25). After 10 years you have $Pe^{r/10}$. This is supposed to be $2P$, so

$$Pe^{r/10} = 2P \implies e^{r/10} = 2$$

$$\implies \frac{r}{10} = \log 2 \implies r = 10 \log 2 = 6.93\%$$

(b) Let the interest rate be $r\%$ per year. If you start with $\$P$, then after t years, you have $P\left(1 + \frac{r}{100 \times 12}\right)^{12t}$, under monthly compounding. This was (2.4.23). After 10 years you have $P\left(1 + \frac{r}{100 \times 12}\right)^{120}$. This is supposed to be $2P$, so

$$P\left(1 + \frac{r}{100 \times 12}\right)^{120} = 2P \implies \left(1 + \frac{r}{1200}\right)^{120} = 2$$

$$\implies 1 + \frac{r}{1200} = 2^{1/120}$$

$$\implies \frac{r}{1200} = 2^{1/120} - 1$$

$$\implies r = 1200(2^{1/120} - 1) = 6.95\%$$

✓ Example 2.4.28 Pension planning

A 25 year old graduate of UBC is given $\$50,000$ which is invested at 5% per year compounded continuously. The graduate also intends to deposit money continuously at the rate of $\$2000$ per year.

1. Find a differential equation that $A(t)$ obeys, assuming that the interest rate remains 5% .
2. Determine the amount of money in the account when the graduate is 65.
3. At age 65, the graduate will start withdrawing money continuously at the rate of W dollars per year. If the money must last until the person is 85, what is the largest possible value of W ?

Solution: (a) Let's consider what happens to A over a very short time interval from time t to time $t + \Delta t$. At time t the account balance is $A(t)$. During the (really short) specified time interval the balance remains very close to $A(t)$ and so earns interest of $\frac{5}{100} \times \Delta t \times A(t)$. During the same time interval, the graduate also deposits an additional $\$2000\Delta t$. So

$$A(t + \Delta t) \approx A(t) + 0.05A(t)\Delta t + 2000\Delta t$$

$$\implies \frac{A(t + \Delta t) - A(t)}{\Delta t} \approx 0.05A(t) + 2000$$

In the limit $\Delta t \rightarrow 0$, the approximation becomes exact and we get

$$\frac{dA}{dt} = 0.05A + 2000$$

(b) The amount of money at time t obeys

$$\frac{dA}{dt} = 0.05A(t) + 2,000 = 0.05(A(t) + 40,000)$$

So by Theorem 2.4.4 (with $a = 0.05$ and $b = -40,000$),

$$A(t) = (A(0) + 40,000)e^{0.05t} - 40,000$$

At time 0 (when the graduate is 25), $A(0) = 50,000$, so the amount of money at time t is

$$A(t) = 90,000 e^{0.05t} - 40,000$$

In particular, when the graduate is 65 years old, $t = 40$ and

$$A(40) = 90,000 e^{0.05 \times 40} - 40,000 = \$625,015.05$$

(c) When the graduate stops depositing money and instead starts withdrawing money at a rate W , the equation for A becomes

$$\frac{dA}{dt} = 0.05A - W = 0.05(A - 20W)$$

assuming that the interest rate remains 5% . This time, Theorem 2.4.4 (with $a = 0.05$ and $b = 20W$) gives

$$A(t) = (A(0) - 20W)e^{0.05t} + 20W$$

If we now reset our clock so that $t = 0$ when the graduate is 65, $A(0) = 625,015.05$. So the amount of money at time t is

$$A(t) = 20W + e^{0.05t}(625,015.05 - 20W)$$

We want the account to be depleted when the graduate is 85. So, we want $A(20) = 0$. This is the case if

$$\begin{aligned} 20W + e^{0.05 \times 20}(625,015.05 - 20W) &= 0 \\ \implies 20W + e(625,015.05 - 20W) &= 0 \\ \implies 20(e - 1)W &= 625,015.05e \\ \implies W &= \frac{625,015.05e}{20(e - 1)} = \$49,437.96 \end{aligned}$$

Exercises

Recall that we are using $\log x$ to denote the logarithm of x with base e . In other courses it is often denoted $\ln x$.

Stage 1

? 1

Below are pairs of functions $y = f(x)$ and differential equations. For each pair, decide whether the function is a solution of the differential equation.

	function	differential equation
(a)	$y = 5(e^x - 3x^2 - 6x - 6)$	$\frac{dy}{dx} = y + 15x^2$
(b)	$y = \frac{-2}{x^2 + 1}$	$y'(x) = xy^2$
(c)	$y = x^{3/2} + x$	$\left(\frac{dy}{dx}\right)^2 + \frac{dy}{dx} = y$

? 2

Following Definition 2.4.1, a separable differential equation has the form

$$\frac{dy}{dx}(x) = f(x)g(y(x)).$$

Show that each of the following equations can be written in this form, identifying $f(x)$ and $g(y)$.

- $3y \frac{dy}{dx} = x \sin y$
- $\frac{dy}{dx} = e^{x+y}$
- $\frac{dy}{dx} + 1 = x$
- $\left(\frac{dy}{dx}\right)^2 - 2x \frac{dy}{dx} + x^2 = 0$

? 3

Suppose we have the following functions:

- y is a differentiable function of x
- f is a function of x , with $\int f(x) dx = F(x)$
- g is a nonzero function of y , with $\int \frac{1}{g(y)} dy = G(y) = G(y(x))$.

In the work below, we set up a solution to the separable differential equation

$$\frac{dy}{dx} = f(x)g(y) = f(x)g(y(x))$$

without using the mnemonic of Equation (*).

By deleting some portion of our work, we can create the solution as it would look using the mnemonic. What portion can be deleted?

Remark: the purpose of this exercise is to illuminate what, exactly, the mnemonic is a shortcut for. Despite its peculiar look, it agrees with what we already know about integration.

$$\frac{dy}{dx} = f(x)g(y(x))$$

Since $g(y(x))$ is a nonzero function, we can divide both sides by it.

$$\frac{1}{g(y(x))} \cdot \frac{dy}{dx} = f(x)$$

If these functions of x are the same, then they have the same antiderivative with respect to x .

$$\int \frac{1}{g(y(x))} \cdot \frac{dy}{dx} dx = \int f(x) dx$$

The left integral is in the correct form for a change of variables to y . To make this easier to see, we'll use a u -substitution, since it's a little more familiar than a y -substitution. If $u = y$, then $\frac{du}{dx} = \frac{dy}{dx}$, so $du = \frac{dy}{dx} dx$.

$$\int \frac{1}{g(u)} du = \int f(x) dx$$

Since u was just the same as y , again for cosmetic reasons, we can swap it back. (Formally, you could have skipped the step above--we just included it to be extra clear that we're not using any integration techniques we haven't seen before.)

$$\int \frac{1}{g(y)} dy = \int f(x) dx$$

We're given the antiderivatives in question.

$$\begin{aligned} G(y) + C_1 &= F(x) + C_2 \\ G(y) &= F(x) + (C_2 - C_1) \end{aligned}$$

where C_1 and C_2 are arbitrary constants. Then also $C_2 - C_1$ is an arbitrary constant, so we might as well call it C .

$$G(y) = F(x) + C$$

? 4

Suppose $y = f(x)$ is a solution to the differential equation $\frac{dy}{dx} = xy$.

True or false: $f(x) + C$ is also a solution, for any constant C .

? 5

Suppose a function $y = f(x)$ satisfies $|y| = Cx$, for some constant $C > 0$.

1. What is the largest possible domain of $f(x)$, given the information at hand?
2. Give an example of function $y = f(x)$ with the following properties, or show that none exists:
 - o $|y| = Cx$,
 - o $\frac{dy}{dx}$ exists for all $x > 0$, and
 - o $y > 0$ for some values of x , and $y < 0$ for others.

? 6

Express the following sentence¹² as a differential equation. You don't have to solve the equation.

About 0.3 percent of the total quantity of morphine in the bloodstream is eliminated every minute.

? 7

Suppose a particular change is occurring in a language, from an old form to a new form.¹³ Let $p(t)$ be the proportion (measured as a number between 0, meaning none, and 1, meaning all) of the time that speakers use the new form. Piotrowski's law¹⁴ predicts the following.

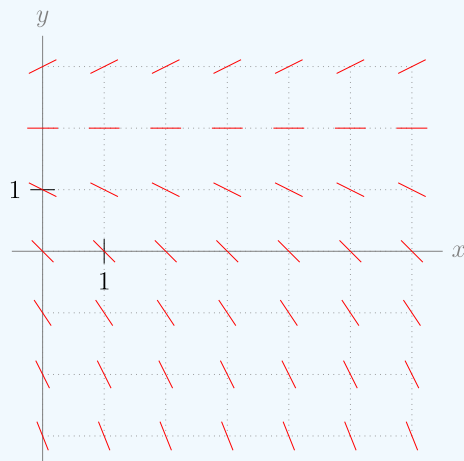
Use of the new form over time spreads at a rate that is proportional to the product of the proportion of the new form and the proportion of the old form.

Express this as a differential equation. You do not need to solve the differential equation.

? 8

Consider the differential equation $y' = \frac{y}{2} - 1$.

1. When $y = 0$, what is y' ?
2. When $y = 2$, what is y' ?
3. When $y = 3$, what is y' ?
4. On the axes below, interpret the marks we have made, and use them to sketch a possible solution to the differential equation.



? 9

Consider the differential equation $y' = y - \frac{x}{2}$.

1. If $y(1) = 0$, what is $y'(1)$?
2. If $y(1) = 2$, what is $y'(1)$?
3. If $y(1) = -2$, what is $y'(1)$?
4. Draw a sketch similar to that of Question 8(d) showing the derivatives of y at the points with integer values for x in $[0, 6]$ and y in $[-3, 3]$.
5. Sketch a possible graph of y .

Stage 2

? 10 (*)

Find the solution to the separable initial value problem:

$$\frac{dy}{dx} = \frac{2x}{e^y}, \quad y(0) = \log 2$$

Express your solution explicitly as $y = y(x)$.

? 11 (*)

Find the solution $y(x)$ of $\frac{dy}{dx} = \frac{xy}{x^2 + 1}$, $y(0) = 3$.

? 12 (*)

Solve the differential equation $y'(t) = e^{\frac{y}{3}} \cos t$. You should express the solution $y(t)$ in terms of t explicitly.

? 13 (*)

Solve the differential equation

$$\frac{dy}{dx} = xe^{x^2 - \log(y^2)}$$

? 14 (*)

Let $y = y(x)$. Find the general solution of the differential equation $y' = xe^y$.

? 15 (*)

Find the solution to the differential equation $\frac{yy'}{e^x - 2x} = \frac{1}{y}$ that satisfies $y(0) = 3$. Solve completely for y as a function of x .

? 16 (*)

Find the function $y = f(x)$ that satisfies

$$\frac{dy}{dx} = -xy^3 \quad \text{and} \quad f(0) = -\frac{1}{4}$$

? 17 (*)

Find the function $y = y(x)$ that satisfies $y(1) = 4$ and

$$\frac{dy}{dx} = \frac{15x^2 + 4x + 3}{y}$$

? 18 (*)

Find the solution $y(x)$ of $y' = x^3y$ with $y(0) = 1$.

? 19 (*)

Find the solution of the initial value problem

$$x \frac{dy}{dx} + y = y^2 \quad y(1) = -1$$

? 20 (*)

A function $f(x)$ is always positive, has $f(0) = e$ and satisfies $f'(x) = x f(x)$ for all x . Find this function.

? 21 (*)

Solve the following initial value problem:

$$\frac{dy}{dx} = \frac{1}{(x^2 + x)y} \quad y(1) = 2$$

? 22 (*)

Find the solution of the differential equation $\frac{1 + \sqrt{y^2 - 4}}{\tan x} y' = \frac{\sec x}{y}$ that satisfies $y(0) = 2$. You don't have to solve for y in terms of x .

? 23 (*)

The fish population in a lake is attacked by a disease at time $t = 0$, with the result that the size $P(t)$ of the population at time $t \geq 0$ satisfies

$$\frac{dP}{dt} = -k\sqrt{P}$$

where k is a positive constant. If there were initially 90,000 fish in the lake and 40,000 were left after 6 weeks, when will the fish population be reduced to 10,000?

? 24 (*)

An object of mass m is projected straight upward at time $t = 0$ with initial speed v_0 . While it is going up, the only forces acting on it are gravity (assumed constant) and a drag force proportional to the square of the object's speed $v(t)$. It follows that the differential equation of motion is

$$m \frac{dv}{dt} = -(mg + kv^2)$$

where g and k are positive constants. At what time does the object reach its highest point?

? 25 (*)

A motor boat is traveling with a velocity of 40 ft/sec when its motor shuts off at time $t = 0$. Thereafter, its deceleration due to water resistance is given by

$$\frac{dv}{dt} = -k v^2$$

where k is a positive constant. After 10 seconds, the boat's velocity is 20 ft/sec.

1. What is the value of k ?
2. When will the boat's velocity be 5 ft/sec?

? 26 (*)

Consider the initial value problem $\frac{dx}{dt} = k(3-x)(2-x)$, $x(0) = 1$, where k is a positive constant. (This kind of problem occurs in the analysis of certain chemical reactions.)

1. Solve the initial value problem. That is, find x as a function of t .
2. What value will $x(t)$ approach as t approaches $+\infty$.

? 27 (*)

The quantity $P = P(t)$, which is a function of time t , satisfies the differential equation

$$\frac{dP}{dt} = 4P - P^2$$

and the initial condition $P(0) = 2$.

1. Solve this equation for $P(t)$.
2. What is P when $t = 0.5$? What is the limiting value of P as t becomes large?

? 28 (*)

An object moving in a fluid has an initial velocity v of 400 m/min. The velocity is decreasing at a rate proportional to the square of the velocity. After 1 minute the velocity is 200 m/min.

1. Give a differential equation for the velocity $v = v(t)$ where t is time.
2. Solve this differential equation.
3. When will the object be moving at 50 m/min?

Stage 3

? 29 (*)

An investor places some money in a mutual fund where the interest is compounded continuously and where the interest rate fluctuates between 4% and 8%. Assume that the amount of money $B = B(t)$ in the account in dollars after t years satisfies the differential equation

$$\frac{dB}{dt} = (0.06 + 0.02 \sin t) B$$

1. Solve this differential equation for B as a function of t .
2. If the initial investment is \$1000, what will the balance be at the end of two years?

? 30 (*)

An endowment is an investment account in which the balance ideally remains constant and withdrawals are made on the interest earned by the account. Such an account may be modeled by the initial value problem $B'(t) = aB - m$ for $t \geq 0$, with $B(0) = B_0$. The constant a reflects the annual interest rate, m is the annual rate of withdrawal, and B_0 is the initial balance in the account.

1. Solve the initial value problem with $a = 0.02$ and $B(0) = B_0 = \$30,000$. Note that your answer depends on the constant m .
2. If $a = 0.02$ and $B(0) = B_0 = \$30,000$, what is the annual withdrawal rate m that ensures a constant balance in the account?

? 31 (*)

A certain continuous function $y = y(x)$ satisfies the integral equation

$$y(x) = 3 + \int_0^x (y(t)^2 - 3y(t) + 2) \sin t \, dt$$

for all x in some open interval containing 0. Find $y(x)$ and the largest interval for which (*) holds.

? 32 (*)

A cylindrical water tank, of radius 3 meters and height 6 meters, is full of water when its bottom is punctured. Water drains out through a hole of radius 1 centimeter. If

- $h(t)$ is the height of the water in the tank at time t (in meters) and
- $v(t)$ is the velocity of the escaping water at time t (in meters per second) then
- Torricelli's law states that $v(t) = \sqrt{2gh(t)}$ where $g = 9.8 \text{ m/sec}^2$. Determine how long it takes for the tank to empty.

? 33 (*)

A spherical tank of radius 6 feet is full of mercury when a circular hole of radius 1 inch is opened in the bottom. How long will it take for all of the mercury to drain from the tank?

Use the value $g = 32 \text{ feet/sec}^2$. Also use Torricelli's law, which states when the height of mercury in the tank is h , the speed of the mercury escaping from the tank is $v = \sqrt{2gh}$.

? 34 (*)

Consider the equation

$$f(x) = 3 + \int_0^x (f(t) - 1)(f(t) - 2) \, dt$$

1. What is $f(0)$?
2. Find the differential equation satisfied by $f(x)$.
3. Solve the initial value problem determined in (a) and (b).

? 35 (*)

A tank 2 m tall is to be made with circular cross-sections with radius $r = y^p$. Here y measures the vertical distance from the bottom of the tank and p is a positive constant to be determined. You may assume that when the tank drains, it obeys Torricelli's law, that is

$$A(y) \frac{dy}{dt} = -c\sqrt{y}$$

for some constant c where $A(y)$ is the cross-sectional area of the tank at height y . It is desired that the tank be constructed so that the top half ($y = 2$ to $y = 1$) takes exactly the same amount of time to drain as the bottom half ($y = 1$ to $y = 0$). Determine the value of p so that the tank has this property. Note: it is not possible or necessary to find c for this question.

? 36

Suppose $f(t)$ is a continuous, differentiable function and the root mean square of $f(t)$ on $[a, x]$ is equal to the average of $f(t)$ on $[a, x]$ for all x . That is,

$$\frac{1}{x-a} \int_a^x f(t) dt = \sqrt{\frac{1}{x-a} \int_a^x f^2(t) dt}$$

You may assume $x > a$.

1. Guess a function $f(t)$ for which the average of $f(t)$ is the same as the root mean square of $f(t)$ on any interval.
2. Differentiate both sides of the given equation.
3. Simplify your answer from (b) by using Equation (*) to replace all terms containing $\int_a^x f^2(t) dt$ with terms containing $\int_a^x f(t) dt$.
4. Let $Y(x) = \int_a^x f(t) dt$, so the equation from (c) becomes a differential equation. Find all functions that satisfy it.
5. What is $f(t)$?

? 37

Find the function $y(x)$ such that

$$\frac{d^2}{dx^2} = \frac{2}{y^3} \cdot \frac{dy}{dx}$$

and if $x = -\frac{1}{16} \log 3$, then $y = 1$ and $\frac{dy}{dx} = 3$.

You do not need to solve for y explicitly.

1. Look at the right hand side of the equation. The x -dependence is separated from the y -dependence. That's the reason for the name "separable".
2. Willard Libby, of Chicago University was awarded the Nobel Prize in Chemistry in 1960, for developing radiocarbon dating.
3. The precise transition is $^{14}\text{C} \rightarrow ^{14}\text{N} + e^- + \bar{\nu}_e$ where e^- is an electron and $\bar{\nu}_e$ is an electron neutrino.
4. The British Museum has an Egyptian mathematical text from the seventeenth century B.C.
5. It does not work so well when the object is of a similar size to its surroundings since the temperature of the surroundings will rise as the object cools. It also fails when there are phase transitions involved — for example, an ice-cube melting in a warm room does not obey Newton's law of cooling.
6. This is named after Rev. Thomas Robert Malthus. He described this model in a 1798 paper called "An essay on the principle of population".
7. That is, the population has plenty of food and space to grow.
8. The *2015 Revision of World Population*, a publication of the United Nations, predicts that the world's population in 2100 will be about 11 billion. But "about" covers a pretty large range. They give an 80% confidence interval running from 10 billion to 12.5 billion.
9. The world population really reached 6 billion in about 1999.
10. No pun intended (sorry).
11. There are banks that advertise continuous compounding. You can find some by googling "interest is compounded continuously and paid"
12. The sentence is paraphrased from the Pharmacokinetics website of Université de Lausanne, [Elimination Kinetics](#). The half-life of morphine is given on [the same website](#). Accessed 12 August 2017.
13. An example is the change in German from "wollt" to "wollst" for the second-person conjugation of the verb "wollen." This example is provided by the site *Laws in Quantitative Linguistics*, [Change in Language](#), accessed 18 August 2017.

14. Piotrowski's law is paraphrased from the page Piotrowski-Gesetz on Glottopedia, accessed 18 August 2017. According to this source, the law was based on work by the married couple R. G. Piotrowski and A. A. Piotrowskaja, later generalized by G. Altmann.

This page titled [2.4: Separable Differential Equations](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

3: Sequence and series

You have probably learned about Taylor polynomials¹ and, in particular, that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + E_n(x)$$

where $E_n(x)$ is the error introduced when you approximate e^x by its Taylor polynomial of degree n . You may have even seen a formula for $E_n(x)$. We are now going to ask what happens as n goes to infinity? Does the error go to zero, giving an exact formula for e^x ? We shall later see that it does and that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

At this point we haven't defined, or developed any understanding of, this infinite sum. How do we compute the sum of an infinite number of terms? Indeed, when does a sum of an infinite number of terms even make sense? Clearly we need to build up foundations to deal with these ideas. Along the way we shall also see other functions for which the corresponding error obeys $\lim_{n \rightarrow \infty} E_n(x) = 0$ for some values of x and not for other values of x .

To motivate the next section, consider using the above formula with $x = 1$ to compute the number e :

$$e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}$$

As we stated above, we don't yet understand what to make of this infinite number of terms, but we might try to sneak up on it by thinking about what happens as we take more and more terms.

1 term	$1 = 1$
2 terms	$1 + 1 = 2$
3 terms	$1 + 1 + \frac{1}{2} = 2.5$
4 terms	$1 + 1 + \frac{1}{2} + \frac{1}{6} = 2.666666 \dots$
5 terms	$1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} = 2.708333 \dots$
6 terms	$1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} = 2.716666 \dots$

By looking at the infinite sum in this way, we naturally obtain a sequence of numbers

$$\{ 1, 2, 2.5, 2.666666, \dots, 2.708333, \dots, 2.716666, \dots, \dots \}.$$

The key to understanding the original infinite sum is to understand the behaviour of this sequence of numbers — in particular, what do the numbers do as we go further and further? Does it settle down² to a given limit?

1. Now would be an excellent time to quickly read over your notes on the topic.
2. You will notice a great deal of similarity between the results of the next section and “limits at infinity” which was covered last term.

- [3.1: Sequences](#)
- [3.2: Series](#)
- [3.3: Convergence Tests](#)
- [3.4: Absolute and Conditional Convergence](#)
- [3.5: Power Series](#)

[3.6: Taylor Series](#)

[3.7: Optional — Rational and irrational numbers](#)

This page titled [3: Sequence and series](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

3.1: Sequences

In the discussion above we used the term “sequence” without giving it a precise mathematical meaning. Let us rectify this now.

Definition 3.1.1

A sequence is a list of infinitely ¹ many numbers with a specified order. It is denoted

$$\{a_1, a_2, a_3, \dots, a_n, \dots\} \quad \text{or} \quad \{a_n\} \quad \text{or} \quad \{a_n\}_{n=1}^{\infty}$$

We will often specify a sequence by writing it more explicitly, like

$$\{a_n = f(n)\}_{n=1}^{\infty}$$

where $f(n)$ is some function from the natural numbers to the real numbers.

Example 3.1.2 Three sequences and another one

Here are three sequences.

$$\begin{array}{ll} \left\{1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots\right\} & \text{or} \quad \left\{a_n = \frac{1}{n}\right\}_{n=1}^{\infty} \\ \left\{1, 2, 3, \dots, n, \dots\right\} & \text{or} \quad \left\{a_n = n\right\}_{n=1}^{\infty} \\ \left\{1, -1, 1, -1, \dots, (-1)^{n-1}, \dots\right\} & \text{or} \quad \left\{a_n = (-1)^{n-1}\right\}_{n=1}^{\infty} \end{array}$$

It is not necessary that there be a simple explicit formula for the n^{th} term of a sequence. For example the decimal digits of π is a perfectly good sequence

$$\{3, 1, 4, 1, 5, 9, 2, 6, 5, 3, 5, 8, 9, 7, 9, 3, 2, 3, 8, 4, 6, 2, 6, 4, \dots\}$$

but there is no simple formula² for the n^{th} digit.

Our primary concern with sequences will be the behaviour of a_n as n tends to infinity and, in particular, whether or not a_n “settles down” to some value as n tends to infinity.

Definition 3.1.3

A sequence $\{a_n\}_{n=1}^{\infty}$ is said to converge to the limit A if a_n approaches A as n tends to infinity. If so, we write

$$\lim_{n \rightarrow \infty} a_n = A \quad \text{or} \quad a_n \rightarrow A \text{ as } n \rightarrow \infty$$

A sequence is said to converge if it converges to some limit. Otherwise it is said to diverge.

The reader should immediately recognise the similarity with limits at infinity

$$\lim_{x \rightarrow \infty} f(x) = L \quad \text{if} \quad f(x) \rightarrow L \text{ as } x \rightarrow \infty$$

Example 3.1.4 Convergence in Example 3.1.2

Three of the four sequences in Example 3.1.2 diverge:

- The sequence $\{a_n = n\}_{n=1}^{\infty}$ diverges because a_n grows without bound, rather than approaching some finite value, as n tends to infinity.
- The sequence $\{a_n = (-1)^{n-1}\}_{n=1}^{\infty}$ diverges because a_n oscillates between $+1$ and -1 rather than approaching a single value as n tends to infinity.
- The sequence of the decimal digits of π also diverges, though the proof that this is the case is a bit beyond us right now³.

The other sequence in Example 3.1.2 has $a_n = \frac{1}{n}$. As n tends to infinity, $\frac{1}{n}$ tends to zero. So

$$\lim_{n \rightarrow \infty} \frac{1}{n} = 0$$

✓ Example 3.1.5 $\lim_{n \rightarrow \infty} \frac{n}{2n+1}$

Here is a little less trivial example. To study the behaviour of $\frac{n}{2n+1}$ as $n \rightarrow \infty$, it is a good idea to write it as

$$\frac{n}{2n+1} = \frac{1}{2 + \frac{1}{n}}$$

As $n \rightarrow \infty$, the $\frac{1}{n}$ in the denominator tends to zero, so that the denominator $2 + \frac{1}{n}$ tends to 2 and $\frac{1}{2 + \frac{1}{n}}$ tends to $\frac{1}{2}$. So

$$\lim_{n \rightarrow \infty} \frac{n}{2n+1} = \lim_{n \rightarrow \infty} \frac{1}{2 + \frac{1}{n}} = \frac{1}{2}$$

Notice that in this last example, we are really using techniques that we used before to study infinite limits like $\lim_{x \rightarrow \infty} f(x)$. This experience can be easily transferred to dealing with $\lim_{n \rightarrow \infty} a_n$ limits by using the following result.

🔗 Theorem 3.1.6

If

$$\lim_{x \rightarrow \infty} f(x) = L$$

and if $a_n = f(n)$ for all positive integers n , then

$$\lim_{n \rightarrow \infty} a_n = L$$

✓ Example 3.1.7 $\lim_{n \rightarrow \infty} e^{-n}$

Set $f(x) = e^{-x}$. Then $e^{-n} = f(n)$ and

$$\text{since } \lim_{x \rightarrow \infty} e^{-x} = 0 \quad \text{we know that} \quad \lim_{n \rightarrow \infty} e^{-n} = 0$$

The bulk of the rules for the arithmetic of limits of functions that you already know also apply to the limits of sequences. That is, the rules you learned to work with limits such as $\lim_{x \rightarrow \infty} f(x)$ also apply to limits like $\lim_{n \rightarrow \infty} a_n$.

🔗 Theorem 3.1.8 Arithmetic of limits

Let A , B and C be real numbers and let the two sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ converge to A and B respectively. That is, assume that

$$\lim_{n \rightarrow \infty} a_n = A \quad \lim_{n \rightarrow \infty} b_n = B$$

Then the following limits hold.

- $\lim_{n \rightarrow \infty} [a_n + b_n] = A + B$

(The limit of the sum is the sum of the limits.)

- $\lim_{n \rightarrow \infty} [a_n - b_n] = A - B$

(The limit of the difference is the difference of the limits.)

- $\lim_{n \rightarrow \infty} C a_n = C A$.

$$4. \lim_{n \rightarrow \infty} a_n b_n = AB$$

(The limit of the product is the product of the limits.)

$$5. \text{ If } B \neq 0 \text{ then } \lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{A}{B}$$

(The limit of the quotient is the quotient of the limits **provided** the limit of the denominator is not zero.)

We use these rules to evaluate limits of more complicated sequences in terms of the limits of simpler sequences — just as we did for limits of functions.

✓ Example 3.1.9 Arithmetic of limits

Combining Examples 3.1.5 and 3.1.7,

$$\begin{aligned} \lim_{n \rightarrow \infty} \left[\frac{n}{2n+1} + 7e^{-n} \right] &= \lim_{n \rightarrow \infty} \frac{n}{2n+1} + \lim_{n \rightarrow \infty} 7e^{-n} && \text{by Theorem 3.1.8(a)} \\ &= \lim_{n \rightarrow \infty} \frac{n}{2n+1} + 7 \lim_{n \rightarrow \infty} e^{-n} && \text{by Theorem 3.1.8(c)} \end{aligned}$$

and then using Examples 3.1.5 and 3.1.7

$$\begin{aligned} &= \frac{1}{2} + 7 \cdot 0 \\ &= \frac{1}{2} \end{aligned}$$

There is also a squeeze theorem for sequences.

🔗 Theorem 3.1.10 Squeeze theorem

If $a_n \leq c_n \leq b_n$ for all natural numbers n , and if

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = L$$

then

$$\lim_{n \rightarrow \infty} c_n = L$$

✓ Example 3.1.11 A simple squeeze

In this example we use the squeeze theorem to evaluate

$$\lim_{n \rightarrow \infty} \left[1 + \frac{\pi_n}{n} \right]$$

where π_n is the n^{th} decimal digit of π . That is,

$$\pi_1 = 3 \quad \pi_2 = 1 \quad \pi_3 = 4 \quad \pi_4 = 1 \quad \pi_5 = 5 \quad \pi_6 = 9 \quad \dots$$

We do not have a simple formula for π_n . But we do know that

$$0 \leq \pi_n \leq 9 \implies 0 \leq \frac{\pi_n}{n} \leq \frac{9}{n} \implies 1 \leq 1 + \frac{\pi_n}{n} \leq 1 + \frac{9}{n}$$

and we also know that

$$\lim_{n \rightarrow \infty} 1 = 1 \quad \lim_{n \rightarrow \infty} \left[1 + \frac{9}{n} \right] = 1$$

So the squeeze theorem with $a_n = 1$, $b_n = 1 + \frac{9}{n}$, and $c_n = 1 + \frac{\pi_n}{n}$ gives

$$\lim_{n \rightarrow \infty} \left[1 + \frac{\pi_n}{n} \right] = 1$$

Finally, recall that we can compute the limit of the composition of two functions using continuity. In the same way, we have the following result:

Theorem 3.1.12 Continuous functions of limits

If $\lim_{n \rightarrow \infty} a_n = L$ and if the function $g(x)$ is continuous at L , then

$$\lim_{n \rightarrow \infty} g(a_n) = g(L)$$

Example 3.1.13 $\lim_{n \rightarrow \infty} \sin \frac{\pi n}{2n+1}$

Write $\sin \frac{\pi n}{2n+1} = g\left(\frac{n}{2n+1}\right)$ with $g(x) = \sin(\pi x)$. We saw, in Example 3.1.5 that

$$\lim_{n \rightarrow \infty} \frac{n}{2n+1} = \frac{1}{2}$$

Since $g(x) = \sin(\pi x)$ is continuous at $x = \frac{1}{2}$, which is the limit of $\frac{n}{2n+1}$, we have

$$\lim_{n \rightarrow \infty} \sin \frac{\pi n}{2n+1} = \lim_{n \rightarrow \infty} g\left(\frac{n}{2n+1}\right) = g\left(\frac{1}{2}\right) = \sin \frac{\pi}{2} = 1$$

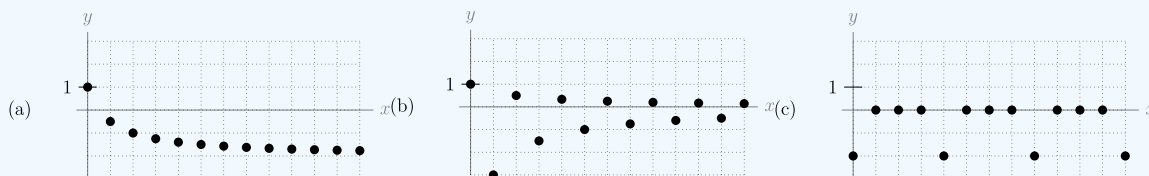
With this introduction to sequences and some tools to determine their limits, we can now return to the problem of understanding infinite sums.

Exercises

Stage 1

1

Assuming the sequences continue as shown, estimate the limit of each sequence from its graph.



2

Suppose a_n and b_n are sequences, and $a_n = b_n$ for all $n \geq 100$, but $a_n \neq b_n$ for $n < 100$.

True or false: $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n$.

3

Let $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$, and $\{c_n\}_{n=1}^{\infty}$, be sequences with $\lim_{n \rightarrow \infty} a_n = A$, $\lim_{n \rightarrow \infty} b_n = B$, and $\lim_{n \rightarrow \infty} c_n = C$. Assume A , B , and C are nonzero real numbers.

Evaluate the limits of the following sequences.

1. $\frac{a_n - b_n}{c_n}$
2. $\frac{c_n}{n}$

3. $\frac{a_{2n+5}}{b_n}$

? 4

Give an example of a sequence $\{a_n\}_{n=1}^{\infty}$ with the following properties:

- $a_n > 1000$ for all $n \leq 1000$,
- $a_{n+1} < a_n$ for all n , and
- $\lim_{n \rightarrow \infty} a_n = -2$

? 5

Give an example of a sequence $\{a_n\}_{n=1}^{\infty}$ with the following properties:

- $a_n > 0$ for all even n ,
- $a_n < 0$ for all odd n ,
- $\lim_{n \rightarrow \infty} a_n$ does not exist.

? 6

Give an example of a sequence $\{a_n\}_{n=1}^{\infty}$ with the following properties:

- $a_n > 0$ for all even n ,
- $a_n < 0$ for all odd n ,
- $\lim_{n \rightarrow \infty} a_n$ exists.

? 7

The limits of the sequences below can be evaluated using the squeeze theorem. For each sequence, choose an upper bounding sequence and lower bounding sequence that will work with the squeeze theorem.

1. $a_n = \frac{\sin n}{n}$

2. $b_n = \frac{n^2}{e^n(7 + \sin n - 5 \cos n)}$

3. $c_n = (-n)^{-n}$

? 8

Below is a list of sequences, and a list of functions.

1. Match each sequence a_n to any and all functions $f(x)$ such that $f(n) = a_n$ for all whole numbers n .

2. Match each sequence a_n to any and all functions $f(x)$ such that $\lim_{n \rightarrow \infty} a_n = \lim_{x \rightarrow \infty} f(x)$.

$$a_n = 1 + \frac{1}{n}$$

$$f(x) = \cos(\pi x)$$

$$b_n = 1 + \frac{1}{|n|}$$

$$g(x) = \frac{\cos(\pi x)}{x}$$

$$c_n = e^{-n}$$

$$h(x) = \begin{cases} \frac{x+1}{x} & x \text{ is a whole number} \\ 1 & \text{else} \end{cases}$$

$$d_n = (-1)^n$$

$$i(x) = \begin{cases} \frac{x+1}{x} & x \text{ is a whole number} \\ 0 & \text{else} \end{cases}$$

$$e_n = \frac{(-1)^n}{n}$$

$$j(x) = \frac{1}{e^x}$$

? 9

Let $\{a_n\}_{n=1}^{\infty}$ be a sequence defined by $a_n = \cos n$.

1. Give three different whole numbers n that are within 0.1 of an odd integer multiple of π , and find the corresponding values of a_n .
2. Give three different whole numbers n such that a_n is close to 0. Justify your answers.

Remark: this demonstrates intuitively, though not rigorously, why $\lim_{n \rightarrow \infty} \cos n$ is undefined. We consistently find terms in the series that are close to -1 , and also consistently find terms in the series that are close to 1. Contrast this to a series like $\{\cos(2\pi n)\}$, whose terms are always 1, and whose limit therefore is 1. It is possible to turn the ideas of this question into a rigorous proof that $\lim_{n \rightarrow \infty} \cos n$ is undefined. See the solution.

Stage 2

? 10

Determine the limits of the following sequences.

1. $a_n = \frac{3n^2 - 2n + 5}{4n + 3}$
2. $b_n = \frac{3n^2 - 2n + 5}{4n^2 + 3}$
3. $c_n = \frac{3n^2 - 2n + 5}{4n^3 + 3}$

? 11

Determine the limit of the sequence $a_n = \frac{4n^3 - 21}{n^e + \frac{1}{n}}$.

? 12

Determine the limit of the sequence $b_n = \frac{\sqrt[4]{n} + 1}{\sqrt{9n + 3}}$.

? 13

Determine the limit of the sequence $c_n = \frac{\cos(n + n^2)}{n}$.

? 14

Determine the limit of the sequence $a_n = \frac{n^{\sin n}}{n^2}$.

? 15

Determine the limit of the sequence $d_n = e^{-1/n}$.

? 16

Determine the limit of the sequence $a_n = \frac{1 + 3 \sin(n^2) - 2 \sin n}{n}$.

? 17

Determine the limit of the sequence $b_n = \frac{e^n}{2^n + n^2}$.

? 18 (*)

Find the limit, if it exists, of the sequence $\{a_k\}$, where

$$a_k = \frac{k! \sin^3 k}{(k+1)!}$$

? 19 (*)

Consider the sequence $\left\{(-1)^n \sin\left(\frac{1}{n}\right)\right\}$. State whether this sequence converges or diverges, and if it converges give its limit.

? 20 (*)

Evaluate $\lim_{n \rightarrow \infty} \left[\frac{6n^2 + 5n}{n^2 + 1} + 3 \cos(1/n^2) \right]$.

Stage 3

? 21 (*)

Find the limit of the sequence $\left\{ \log\left(\sin \frac{1}{n}\right) + \log(2n) \right\}$.

? 22

Evaluate $\lim_{n \rightarrow \infty} \left[\sqrt{n^2 + 5n} - \sqrt{n^2 - 5n} \right]$.

? 23

Evaluate $\lim_{n \rightarrow \infty} \left[\sqrt{n^2 + 5n} - \sqrt{2n^2 - 5} \right]$.

? 24

Evaluate the limit of the sequence $\left\{ n \left[\left(2 + \frac{1}{n}\right)^{100} - 2^{100} \right] \right\}_{n=1}^{\infty}$.

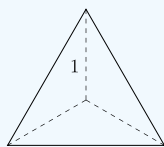
? 25

Write a sequence $\{a_n\}_{n=1}^{\infty}$ whose limit is $f'(a)$ for a function $f(x)$ that is differentiable at the point a .

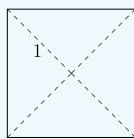
Your answer will depend on f and a .

? 26

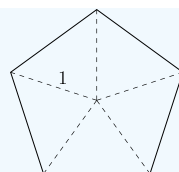
Let $\{A_n\}_{n=3}^{\infty}$ be the area of a regular polygon with n sides, with the distance from the centroid of the polygon to each corner equal to 1.



$$A(3) = \frac{3\sqrt{3}}{4}$$



$$A(4) = 2$$



$$A(5) = 2.5 \sin(0.4\pi)$$

1. By dividing the polygon into n triangles, give a formula for A_n .
2. What is $\lim_{n \rightarrow \infty} A_n$?

? 27

Suppose we define a sequence $\{f_n\}$, which depends on some constant x , as the following:

$$f_n(x) = \begin{cases} 1 & n \leq x < n+1 \\ 0 & \text{else} \end{cases}$$

For a fixed constant $x \geq 1$, $\{f_n\}$ is the sequence $\{0, 0, 0, \dots, 0, 1, 0, \dots, 0, 0, 0, \dots\}$. The sole nonzero element comes in position k , where k is what we get when we round x down to a whole number. If $x < 1$, then the sequence consists of all zeroes.

Since we can plug in different values of x , we can think of $f_n(x)$ as a function of sequences: a different x gives you a different sequence. On the other hand, if we imagine fixing n , then $f_n(x)$ is just a function, where $f_n(x)$ gives the n th term in the sequence corresponding to x .

1. Sketch the curve $y = f_2(x)$.
2. Sketch the curve $y = f_3(x)$.
3. Define $A_n = \int_0^\infty f_n(x) dx$. Give a simple description of the sequence $\{A_n\}_{n=1}^\infty$.
4. Evaluate $\lim_{n \rightarrow \infty} A_n$.
5. Evaluate $\lim_{n \rightarrow \infty} f_n(x)$ for a constant x , and call the result $g(x)$.
6. Evaluate $\int_0^\infty g(x) dx$.

? 28

Determine the limit of the sequence $b_n = \left(1 + \frac{3}{n} + \frac{5}{n^2}\right)^n$.

? 29

A sequence $\{a_n\}_{n=1}^\infty$ of real numbers satisfies the recursion relation $a_{n+1} = \frac{a_n + 8}{3}$ for $n \geq 2$.

1. Suppose $a_1 = 4$. What is $\lim_{n \rightarrow \infty} a_n$?
2. Find x if $x = \frac{x+8}{3}$.
3. Suppose $a_1 = 1$. Show that $\lim_{n \rightarrow \infty} a_n = L$, where L is the solution to equation above.

? 30

Zipf's Law applied to word frequency can be phrased as follows:

The most-used word in a language is used n times as frequently as the n -th most word used in a language.

1. Suppose the sequence $\{w_1, w_2, w_3, \dots\}$ is a list of all words in a language, where w_n is the word that is the n th most frequently used. Let f_n be the frequency of word w_n . Is $\{f_1, f_2, f_3, \dots\}$ an increasing sequence or a decreasing sequence?
2. Give a general formula for f_n , treating f_1 as a constant.
3. Suppose in a language, w_1 (the most frequently used word) has frequency 6%. If the language follows Zipf's Law, then what frequency does w_3 have?
4. Suppose $f_6 = 0.3\%$ for a language following Zipf's law. What is f_{10} ?
5. The word “the” is the most-used word in contemporary American English. In a collection of about 450 million words, “the” appeared 22,038,615 times. The second-most used word is “be,” followed by “and.” About how many usages of these words do you expect in the same collection of 450 million words?

Sources:

- Zipf's word frequency law in natural language: A critical review and future directions, Steven T. Piantadosi. *Psychon Bull Rev.* 2014 Oct; 21(5): 1112–1130. Accessed online 11 October 2017.
- [Word Frequency Data](#). Accessed online 11 October 2017.

1. For the more pedantic reader, here we mean a countably infinite list of numbers. The interested (pedantic or otherwise) reader should look up countable and uncountable sets.
2. There is, however, a remarkable result due to Bailey, Borwein and Plouffe that can be used to compute the n^{th} binary digit of π (i.e. writing π in base 2 rather than base 10) without having to work out the preceding digits.
3. If the digits of π were to converge, then π would have to be a rational number. The irrationality of π (that it cannot be written as a fraction) was first proved by Lambert in 1761. Niven's 1947 proof is more accessible and we invite the interested reader to use their favourite search engine to find step-by-step guides to that proof.

This page titled [3.1: Sequences](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

3.2: Series

A series is a sum

$$a_1 + a_2 + a_3 + \cdots + a_n + \cdots$$

of infinitely many terms. In summation notation, it is written

$$\sum_{n=1}^{\infty} a_n$$

You already have a lot of experience with series, though you might not realise it. When you write a number using its decimal expansion you are really expressing it as a series. Perhaps the simplest example of this is the decimal expansion of $\frac{1}{3}$:

$$\frac{1}{3} = 0.3333 \dots$$

Recall that the expansion written in this way actually means

$$0.333333 \dots = \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \frac{3}{10000} + \cdots = \sum_{n=1}^{\infty} \frac{3}{10^n}$$

The summation index n is of course a dummy index. You can use any symbol you like (within reason) for the summation index.

$$\sum_{n=1}^{\infty} \frac{3}{10^n} = \sum_{i=1}^{\infty} \frac{3}{10^i} = \sum_{j=1}^{\infty} \frac{3}{10^j} = \sum_{\ell=1}^{\infty} \frac{3}{10^\ell}$$

A series can be expressed using summation notation in many different ways. For example the following expressions all represent the same series:

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{3}{10^n} &= \overbrace{\frac{3}{10}}^{n=1} + \overbrace{\frac{3}{100}}^{n=2} + \overbrace{\frac{3}{1000}}^{n=3} + \cdots \\ \sum_{j=2}^{\infty} \frac{3}{10^{j-1}} &= \overbrace{\frac{3}{10}}^{j=2} + \overbrace{\frac{3}{100}}^{j=3} + \overbrace{\frac{3}{1000}}^{j=4} + \cdots \\ \sum_{\ell=0}^{\infty} \frac{3}{10^{\ell+1}} &= \overbrace{\frac{3}{10}}^{\ell=0} + \overbrace{\frac{3}{100}}^{\ell=1} + \overbrace{\frac{3}{1000}}^{\ell=3} + \cdots \\ \frac{3}{10} + \sum_{n=2}^{\infty} \frac{3}{10^n} &= \frac{3}{10} + \overbrace{\frac{3}{100}}^{n=2} + \overbrace{\frac{3}{1000}}^{n=3} + \cdots \end{aligned}$$

We can get from the first line to the second line by substituting $n = j - 1$ — don't forget to also change the limits of summation (so that $n = 1$ becomes $j - 1 = 1$ which is rewritten as $j = 2$). To get from the first line to the third line, substitute $n = \ell + 1$ everywhere, including in the limits of summation (so that $n = 1$ becomes $\ell + 1 = 1$ which is rewritten as $\ell = 0$).

Whenever you are in doubt as to what series a summation notation expression represents, it is a good habit to write out the first few terms, just as we did above.

Of course, at this point, it is not clear whether the sum of infinitely many terms adds up to a finite number or not. In order to make sense of this we will recast the problem in terms of the convergence of sequences (hence the discussion of the previous section). Before we proceed more formally let us illustrate the basic idea with a few simple examples.

✓ Example 3.2.1 $\sum_{n=1}^{\infty} \frac{3}{10^n}$

As we have just seen above the series $\sum_{n=1}^{\infty} \frac{3}{10^n}$ is

$$\sum_{n=1}^{\infty} \frac{3}{10^n} = \overbrace{\frac{3}{10}}^{n=1} + \overbrace{\frac{3}{100}}^{n=2} + \overbrace{\frac{3}{1000}}^{n=3} + \dots$$

Notice that the n^{th} term in that sum is

$$3 \times 10^{-n} = \overbrace{0.00 \dots 03}^{n-1 \text{ zeroes}}$$

So the sum of the first 5, 10, 15 and 20 terms in that series are

$$\begin{aligned} \sum_{n=1}^5 \frac{3}{10^n} &= 0.33333 & \sum_{n=1}^{10} \frac{3}{10^n} &= 0.3333333333 \\ \sum_{n=1}^{15} \frac{3}{10^n} &= 0.333333333333333 & \sum_{n=1}^{20} \frac{3}{10^n} &= 0.33333333333333333333 \end{aligned}$$

It sure looks like that, as we add more and more terms, we get closer and closer to $0.\dot{3} = \frac{1}{3}$. So it is very reasonable¹ to define $\sum_{n=1}^{\infty} \frac{3}{10^n}$ to be $\frac{1}{3}$.

✓ Example 3.2.2 $\sum_{n=1}^{\infty} 1$ and $\sum_{n=1}^{\infty} (-1)^n$

Every term in the series $\sum_{n=1}^{\infty} 1$ is exactly 1. So the sum of the first N terms is exactly N . As we add more and more terms this grows unboundedly. So it is very reasonable to say that the series $\sum_{n=1}^{\infty} 1$ diverges.

The series

$$\sum_{n=1}^{\infty} (-1)^n = \overbrace{(-1)}^{n=1} + \overbrace{1}^{n=2} + \overbrace{(-1)}^{n=3} + \overbrace{1}^{n=4} + \overbrace{(-1)}^{n=5} + \dots$$

So the sum of the first N terms is 0 if N is even and -1 if N is odd. As we add more and more terms from the series, the sum alternates between 0 and -1 for ever and ever. So the sum of all infinitely many terms does not make any sense and it is again reasonable to say that the series $\sum_{n=1}^{\infty} (-1)^n$ diverges.

In the above examples we have tried to understand the series by examining the sum of the first few terms and then extrapolating as we add in more and more terms. That is, we tried to sneak up on the infinite sum by looking at the limit of (partial) sums of the first few terms. This approach can be made into a more formal rigorous definition. More precisely, to define what is meant by the infinite sum $\sum_{n=1}^{\infty} a_n$, we approximate it by the sum of its first N terms and then take the limit as N tends to infinity.

 Definition 3.2.3

The N^{th} partial sum of the series $\sum_{n=1}^{\infty} a_n$ is the sum of its first N terms

$$S_N = \sum_{n=1}^N a_n.$$

The partial sums form a sequence $\{S_N\}_{N=1}^{\infty}$. If this sequence of partial sums converges $S_N \rightarrow S$ as $N \rightarrow \infty$ then we say that the series $\sum_{n=1}^{\infty} a_n$ converges to S and we write

$$\sum_{n=1}^{\infty} a_n = S$$

If the sequence of partial sums diverges, we say that the series diverges.

✓ Example 3.2.4 Geometric Series

Let a and r be any two fixed real numbers with $a \neq 0$. The series

$$a + ar + ar^2 + \cdots + ar^n + \cdots = \sum_{n=0}^{\infty} ar^n$$

is called the geometric series with first term a and ratio r .

Notice that we have chosen to start the summation index at $n = 0$. That's fine. The first² term is the $n = 0$ term, which is $ar^0 = a$. The second term is the $n = 1$ term, which is $ar^1 = ar$. And so on. We could have also written the series $\sum_{n=1}^{\infty} ar^{n-1}$. That's exactly the same series — the first term is $ar^{n-1}|_{n=1} = ar^{1-1} = a$, the second term is $ar^{n-1}|_{n=2} = ar^{2-1} = ar$, and so on³. Regardless of how we write the geometric series, a is the first term and r is the ratio between successive terms.

Geometric series have the extremely useful property that there is a very simple formula for their partial sums. Denote the partial sum by

$$S_N = \sum_{n=0}^N ar^n = a + ar + ar^2 + \cdots + ar^N.$$

The secret to evaluating this sum is to see what happens when we multiply it by r :

$$\begin{aligned} rS_N &= r(a + ar + ar^2 + \cdots + ar^N) \\ &= ar + ar^2 + ar^3 + \cdots + ar^{N+1} \end{aligned}$$

Notice that this is almost the same⁴ as S_N . The only differences are that the first term, a , is missing and one additional term, ar^{N+1} , has been tacked on the end. So

$$\begin{aligned} S_N &= a + ar + ar^2 + \cdots + ar^N \\ rS_N &= ar + ar^2 + \cdots + ar^N + ar^{N+1} \end{aligned}$$

Hence taking the difference of these expressions cancels almost all the terms:

$$(1 - r)S_N = a - ar^{N+1} = a(1 - r^{N+1})$$

Provided $r \neq 1$ we can divide both side by $1 - r$ to isolate S_N :

$$S_N = a \cdot \frac{1 - r^{N+1}}{1 - r}.$$

On the other hand, if $r = 1$, then

$$S_N = \underbrace{a + a + \cdots + a}_{N+1 \text{ terms}} = a(N + 1)$$

So in summary:

$$S_N = \begin{cases} a \frac{1 - r^{N+1}}{1 - r} & \text{if } r \neq 1 \\ a(N + 1) & \text{if } r = 1 \end{cases}$$

Now that we have this expression we can determine whether or not the series converges. If $|r| < 1$, then r^{N+1} tends to zero as $N \rightarrow \infty$, so that S_N converges to $\frac{a}{1-r}$ as $N \rightarrow \infty$ and

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r} \text{ provided } |r| < 1.$$

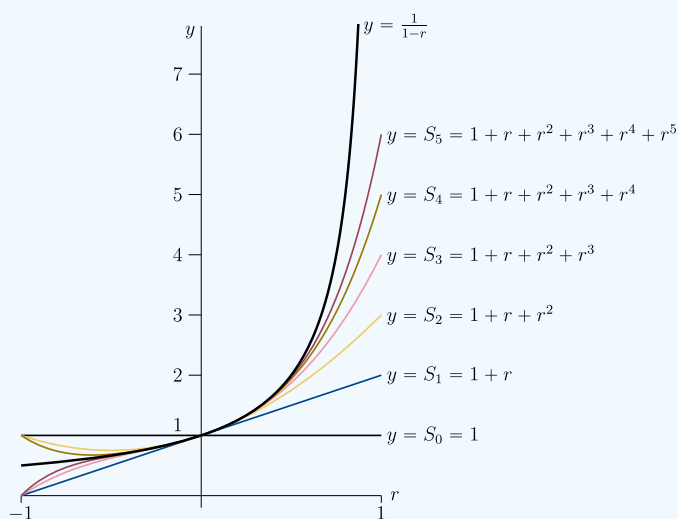
On the other hand if $|r| \geq 1$, S_N diverges. To understand this divergence, consider the following 4 cases:

- If $r > 1$, then r^N grows to ∞ as $N \rightarrow \infty$.
- If $r < -1$, then the magnitude of r^N grows to ∞ , and the sign of r^N oscillates between $+$ and $-$, as $N \rightarrow \infty$.

- If $r = +1$, then $N + 1$ grows to ∞ as $N \rightarrow \infty$.
- If $r = -1$, then r^N just oscillates between $+1$ and -1 as $N \rightarrow \infty$.

In each case the sequence of partial sums does not converge and so the series does not converge.

Here are some sketches of the graphs of $\frac{1}{1-r}$ and S_N , $0 \leq N \leq 5$, for $a = 1$ and $-1 \leq r < 1$.



In these sketches we see that

- when $0 < r < 1$, and also when $-1 < r < 0$ with N odd, we have $S_N = \frac{1-r^{N+1}}{1-r} < \frac{1}{1-r}$. On the other hand, when $-1 < r < 0$ with N even, we have $S_N = \frac{1-r^{N+1}}{1-r} > \frac{1}{1-r}$.
- When $0 < |r| < 1$, $S_N = \frac{1-r^{N+1}}{1-r}$ gets closer and closer to $\frac{1}{1-r}$ as N increases.
- When $r = -1$, S_N just alternates between 0, when N is odd, and 1, when N is even.

We should summarise the results in the previous example in a lemma.

Lemma 3.2.5 Geometric series

Let a and r be real numbers and let $N \geq 0$ be an integer then

$$\sum_{n=0}^N ar^n = \begin{cases} a \frac{1-r^{N+1}}{1-r} & \text{if } r \neq 1 \\ a(N+1) & \text{if } r = 1 \end{cases}.$$

Further, if $|r| < 1$ then

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}.$$

Now that we know how to handle geometric series let's return to Example 3.2.1.

Example 3.2.6 Decimal Expansions

The decimal expansion

$$0.3333 \dots = \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \frac{3}{10000} + \dots = \sum_{n=1}^{\infty} \frac{3}{10^n}$$

is a geometric series with the first term $a = \frac{3}{10}$ and the ratio $r = \frac{1}{10}$. So, by Lemma 3.2.5,

$$0.3333\dots = \sum_{n=1}^{\infty} \frac{3}{10^n} = \frac{\frac{3}{10}}{1 - \frac{1}{10}} = \frac{\frac{3}{10}}{\frac{9}{10}} = \frac{1}{3}$$

just as we would have expected.

We can push this idea further. Consider the repeating decimal expansion:

$$0.16161616\dots = \frac{16}{100} + \frac{16}{10000} + \frac{16}{1000000} + \dots$$

This is another geometric series with the first term $a = \frac{16}{100}$ and the ratio $r = \frac{1}{100}$. So, by Lemma 3.2.5,

$$0.16161616\dots = \sum_{n=1}^{\infty} \frac{16}{100^n} = \frac{\frac{16}{100}}{1 - \frac{1}{100}} = \frac{\frac{16}{100}}{\frac{99}{100}} = \frac{16}{99}$$

again, as expected. In this way any periodic decimal expansion converges to a ratio of two integers — that is, to a rational number⁵.

Here is another more complicated example.

$$\begin{aligned} 0.1234343434\dots &= \frac{12}{100} + \frac{34}{10000} + \frac{34}{1000000} + \dots \\ &= \frac{12}{100} + \sum_{n=2}^{\infty} \frac{34}{100^n} \end{aligned}$$

Now apply Lemma 3.2.5 with $a = \frac{34}{100^2}$ and $r = \frac{1}{100}$

$$\begin{aligned} &= \frac{12}{100} + \frac{34}{10000} \frac{1}{1 - \frac{1}{100}} \\ &= \frac{12}{100} + \frac{34}{10000} \frac{100}{99} \\ &= \frac{1222}{9900} \end{aligned}$$

Typically, it is quite difficult to write down a neat closed form expression for the partial sums of a series. Geometric series are very notable exceptions to this. Another family of series for which we can write down partial sums is called “telescoping series”. These series have the desirable property that many of the terms in the sum cancel each other out rendering the partial sums quite simple.

✓ Example 3.2.7 Telescoping Series

In this example, we are going to study the series $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$. This is a rather artificial series⁶ that has been rigged to illustrate a phenomenon called “telescoping”. Notice that the n^{th} term can be rewritten as

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$$

and so we have

$$a_n = b_n - b_{n+1} \quad \text{where } b_n = \frac{1}{n}.$$

Because of this we get big cancellations when we add terms together. This allows us to get a simple formula for the partial sums of this series.

$$\begin{aligned} S_N &= \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \dots + \frac{1}{N \cdot (N+1)} \\ &= \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \dots + \left(\frac{1}{N} - \frac{1}{N+1}\right) \end{aligned}$$

The second term of each bracket exactly cancels the first term of the following bracket. So the sum “telescopes” leaving just

$$S_N = 1 - \frac{1}{N+1}$$

and we can now easily compute

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \lim_{N \rightarrow \infty} S_N = \lim_{N \rightarrow \infty} \left(1 - \frac{1}{N+1}\right) = 1$$

More generally, if we can write

$$a_n = b_n - b_{n+1}$$

for some other known sequence b_n , then the series telescopes and we can compute partial sums using

$$\begin{aligned} \sum_{n=1}^N a_n &= \sum_{n=1}^N (b_n - b_{n+1}) \\ &= \sum_{n=1}^N b_n - \sum_{n=1}^N b_{n+1} \\ &= b_1 - b_{N+1}. \end{aligned}$$

and hence

$$\sum_{n=1}^{\infty} a_n = b_1 - \lim_{N \rightarrow \infty} b_{N+1}$$

provided this limit exists. Often $\lim_{N \rightarrow \infty} b_{N+1} = 0$ and then $\sum_{n=1}^{\infty} a_n = b_1$. But this does not always happen. Here is an example.

✓ Example 3.2.8 A Divergent Telescoping Series

In this example, we are going to study the series $\sum_{n=1}^{\infty} \log\left(1 + \frac{1}{n}\right)$. Let's start by just writing out the first few terms.

$$\begin{aligned} \sum_{n=1}^{\infty} \log\left(1 + \frac{1}{n}\right) &= \overbrace{\log\left(1 + \frac{1}{1}\right)}^{n=1} + \overbrace{\log\left(1 + \frac{1}{2}\right)}^{n=2} + \overbrace{\log\left(1 + \frac{1}{3}\right)}^{n=3} \\ &\quad + \overbrace{\log\left(1 + \frac{1}{4}\right)}^{n=4} + \dots \\ &= \log(2) + \log\left(\frac{3}{2}\right) + \log\left(\frac{4}{3}\right) \\ &\quad + \log\left(\frac{5}{4}\right) + \dots \end{aligned}$$

This is pretty suggestive since

$$\log(2) + \log\left(\frac{3}{2}\right) + \log\left(\frac{4}{3}\right) + \log\left(\frac{5}{4}\right) = \log\left(2 \times \frac{3}{2} \times \frac{4}{3} \times \frac{5}{4}\right) = \log(5)$$

So let's try using this idea to compute the partial sum S_N :

$$\begin{aligned}
 S_N &= \sum_{n=1}^N \log\left(1 + \frac{1}{n}\right) \\
 &= \overbrace{\log\left(1 + \frac{1}{1}\right)}^{n=1} + \overbrace{\log\left(1 + \frac{1}{2}\right)}^{n=2} + \overbrace{\log\left(1 + \frac{1}{3}\right)}^{n=3} + \cdots \\
 &\quad + \overbrace{\log\left(1 + \frac{1}{N-1}\right)}^{n=N-1} + \overbrace{\log\left(1 + \frac{1}{N}\right)}^{n=N} \\
 &= \log(2) + \log\left(\frac{3}{2}\right) + \log\left(\frac{4}{3}\right) + \cdots + \log\left(\frac{N}{N-1}\right) + \log\left(\frac{N+1}{N}\right) \\
 &= \log\left(2 \times \frac{3}{2} \times \frac{4}{3} \times \cdots \times \frac{N}{N-1} \times \frac{N+1}{N}\right) \\
 &= \log(N+1)
 \end{aligned}$$

Uh oh!

$$\lim_{N \rightarrow \infty} S_N = \lim_{N \rightarrow \infty} \log(N+1) = +\infty$$

This telescoping series diverges! There is an important lesson here. Telescoping series *can diverge*. They do not always converge to b_1 .

As was the case for limits, differentiation and antidifferentiation, we can compute more complicated series in terms of simpler ones by understanding how series interact with the usual operations of arithmetic. It is, perhaps, not so surprising that there are simple rules for addition and subtraction of series and for multiplication of a series by a constant. Unfortunately there are no simple general rules for computing products or ratios of series.

Theorem 3.2.9 Arithmetic of series

Let A , B and C be real numbers and let the two series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ converge to S and T respectively. That is, assume that

$$\sum_{n=1}^{\infty} a_n = S \quad \sum_{n=1}^{\infty} b_n = T$$

Then the following hold.

1. $\sum_{n=1}^{\infty} [a_n + b_n] = S + T$ and $\sum_{n=1}^{\infty} [a_n - b_n] = S - T$
2. $\sum_{n=1}^{\infty} C a_n = C S$.

Example 3.2.10 $\sum_{n=1}^{\infty} \left(\frac{1}{7^n} + \frac{2}{n(n+1)}\right)$

As a simple example of how we use the arithmetic of series Theorem 3.2.9, consider

$$\sum_{n=1}^{\infty} \left[\frac{1}{7^n} + \frac{2}{n(n+1)} \right]$$

We recognize that we know how to compute parts of this sum. We know that

$$\sum_{n=1}^{\infty} \frac{1}{7^n} = \frac{\frac{1}{7}}{1 - \frac{1}{7}} = \frac{1}{6}$$

because it is a geometric series (Example 3.2.4 and Lemma 3.2.5) with first term $a = \frac{1}{7}$ and ratio $r = \frac{1}{7}$. And we know that

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$$

by Example 3.2.7. We can now use Theorem 3.2.9 to build the specified “complicated” series out of these two “simple” pieces.

$$\begin{aligned} \sum_{n=1}^{\infty} \left[\frac{1}{7^n} + \frac{2}{n(n+1)} \right] &= \sum_{n=1}^{\infty} \frac{1}{7^n} + \sum_{n=1}^{\infty} \frac{2}{n(n+1)} && \text{by Theorem 3.2.9(a)} \\ &= \sum_{n=1}^{\infty} \frac{1}{7^n} + 2 \sum_{n=1}^{\infty} \frac{1}{n(n+1)} && \text{by Theorem 3.2.9(b)} \\ &= \frac{1}{6} + 2 \cdot 1 = \frac{13}{6} \end{aligned}$$

Exercises

Stage 1

? 1

Write out the first five partial sums corresponding to the series $\sum_{n=1}^{\infty} \frac{1}{n}$.

You don't need to simplify the terms.

? 2

Every student who comes to class brings their instructor cookies, and leaves them on the instructor's desk. Let C_k be the total number of cookies on the instructor's desk after the k th student comes.

If $C_{11} = 20$, and $C_{10} = 17$, how many cookies did the 11th student bring to class?

? 3

Suppose the sequence of partial sums of the series $\sum_{n=1}^{\infty} a_n$ is $\{S_N\} = \left\{ \frac{N}{N+1} \right\}$.

1. What is $\{a_n\}$?
2. What is $\lim_{n \rightarrow \infty} a_n$?
3. Evaluate $\sum_{n=1}^{\infty} a_n$.

? 4

Suppose the sequence of partial sums of the series $\sum_{n=1}^{\infty} a_n$ is $\{S_N\} = \left\{ (-1)^N + \frac{1}{N} \right\}$.

What is $\{a_n\}$?

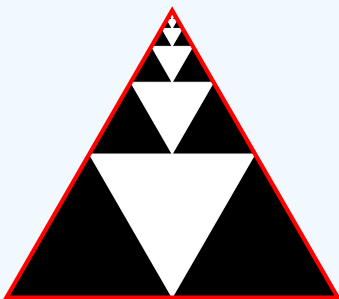
? 5

Let $f(N)$ be a formula for the N th partial sum of $\sum_{n=1}^{\infty} a_n$. (That is, $f(N) = S_N$.) If $f'(N) < 0$ for all $N > 1$, what does that say about a_n ?

Questions 6 through 8 invite you to explore geometric sums in a geometric way. This is complementary to than the algebraic method discussed in the text.

? 6

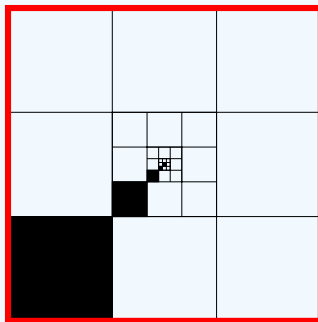
Suppose the triangle outlined in red in the picture below has area one.



1. Express the combined area of the black triangles as a series, assuming the pattern continues forever.
2. Evaluate the series using the picture (*not* the formula from your book).

? 7

Suppose the square outlined in red in the picture below has area one.



1. Express the combined area of the black square as a series, assuming the pattern continues forever.
2. Evaluate the series using the picture (*not* the formula from your book).

? 8

In the style of Questions 6 and 7, draw a picture that represents $\sum_{n=1}^{\infty} \frac{1}{3^n}$ as an area.

? 9

Evaluate $\sum_{n=0}^{100} \frac{1}{5^n}$.

? 10

Every student who comes to class brings their instructor cookies, and leaves them on the instructor's desk. Let C_k be the total number of cookies on the instructor's desk after the k th student comes.

If $C_{20} = 53$, and $C_{10} = 17$, what does $C_{20} - C_{10} = 36$ represent?

? 11

Evaluate $\sum_{n=50}^{100} \frac{1}{5^n}$. (Note the starting index.)

? 12

1. Starting on day $d = 1$, every day you give your friend $\$ \frac{1}{d+1}$, and they give $\$ \frac{1}{d}$ back to you. After a long time, how much money have you gained by this arrangement?

2. Evaluate $\sum_{d=1}^{\infty} \left(\frac{1}{d} - \frac{1}{(d+1)} \right)$.

3. Starting on day $d = 1$, every day your friend gives you $\$(d+1)$, and they take $\$(d+2)$ from you. After a long time, how much money have you gained by this arrangement?

4. Evaluate $\sum_{d=1}^{\infty} ((d+1) - (d+2))$.

? 13

Suppose $\sum_{n=1}^{\infty} a_n = A$, $\sum_{n=1}^{\infty} b_n = B$, and $\sum_{n=1}^{\infty} c_n = C$.

Evaluate $\sum_{n=1}^{\infty} (a_n + b_n + c_{n+1})$.

? 14

Suppose $\sum_{n=1}^{\infty} a_n = A$, $\sum_{n=1}^{\infty} b_n = B \neq 0$, and $\sum_{n=1}^{\infty} c_n = C$.

True or false: $\sum_{n=1}^{\infty} \left(\frac{a_n}{b_n} + c_n \right) = \frac{A}{B} + C$.

Stage 2

? 15 (*)

To what value does the series $1 + \frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \frac{1}{81} + \frac{1}{243} + \dots$ converge?

? 16 (*)

Evaluate $\sum_{k=7}^{\infty} \frac{1}{8^k}$

? 17 (*)

Show that the series $\sum_{k=1}^{\infty} \left(\frac{6}{k^2} - \frac{6}{(k+1)^2} \right)$ converges and find its limit.

? 18 (*)

Find the sum of the convergent series $\sum_{n=3}^{\infty} \left(\cos \left(\frac{\pi}{n} \right) - \cos \left(\frac{\pi}{n+1} \right) \right)$.

? 19 (*)

The n^{th} partial sum of a series $\sum_{n=1}^{\infty} a_n$ is known to have the formula $s_n = \frac{1+3n}{5+4n}$.

- (a) Find an expression for a_n , valid for $n \geq 2$.
- (b) Show that the series $\sum_{n=1}^{\infty} a_n$ converges and find its value.

? 20 (*)

Find the sum of the series $\sum_{n=2}^{\infty} \frac{3 \cdot 4^{n+1}}{8 \cdot 5^n}$. Simplify your answer completely.

? 21 (*)

Relate the number $0.2\bar{3} = 0.233333\dots$ to the sum of a geometric series, and use that to represent it as a rational number (a fraction or combination of fractions, with no decimals).

? 22 (*)

Express $2.656565\dots$ as a rational number, i.e. in the form p/q where p and q are integers.

? 23 (*)

Express the decimal $0.\overline{321} = 0.321321321\dots$ as a fraction.

? 24 (*)

Find the *value* of the convergent series

$$\sum_{n=2}^{\infty} \left(\frac{2^{n+1}}{3^n} + \frac{1}{2n-1} - \frac{1}{2n+1} \right)$$

Simplify your answer completely.

? 25 (*)

Evaluate

$$\sum_{n=1}^{\infty} \left[\left(\frac{1}{3} \right)^n + \left(-\frac{2}{5} \right)^{n-1} \right]$$

? 26 (*)

Find the sum of the series $\sum_{n=0}^{\infty} \frac{1+3^{n+1}}{4^n}$.

? 27

Evaluate $\sum_{n=5}^{\infty} \log \left(\frac{n-3}{n} \right)$.

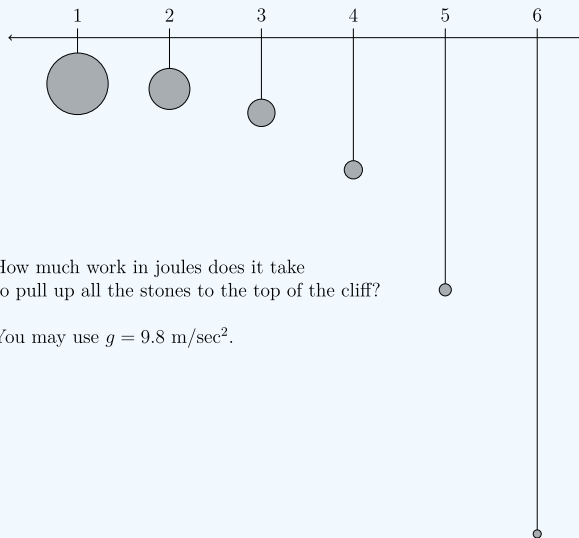
? 28

Evaluate $\sum_{n=2}^{\infty} \left(\frac{2}{n} - \frac{1}{n+1} - \frac{1}{n-1} \right)$.

Stage 3

? 29

An infinitely long, flat cliff has stones hanging off it, attached to thin wire of negligible mass. Starting at position $x = 1$, every metre (at position x , where x is some whole number) the stone has mass $\frac{1}{4^x}$ kg and is hanging 2^x metres below the top of the cliff.



How much work in joules does it take to pull up all the stones to the top of the cliff?

You may use $g = 9.8 \text{ m/sec}^2$.

? 30

Find the combined volume of an infinite collection of spheres, where for each whole number $n = 1, 2, 3, \dots$ there is exactly one sphere of radius $\frac{1}{\pi^n}$.

? 31

Evaluate $\sum_{n=3}^{\infty} \left(\frac{\sin^2 n}{2^n} + \frac{\cos^2(n+1)}{2^{n+1}} \right)$.

? 32

Suppose a series $\sum_{n=1}^{\infty} a_n$ has sequence of partial sums $\{S_N\}$, and the series $\sum_{N=1}^{\infty} S_N$ has sequence of partial sums

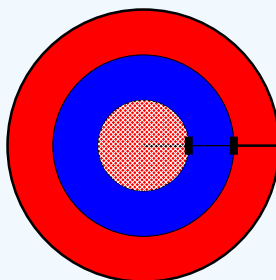
$$\{S_M\} = \left\{ \sum_{N=1}^M S_N \right\}.$$

If $S_M = \frac{M+1}{M}$, what is a_n ?

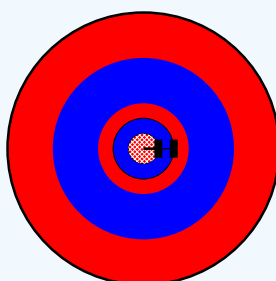
Create a bullseye using the following method:

Starting with a red circle of area 1, divide the radius into thirds, creating two rings and a circle. Colour the middle ring blue.

Continue the pattern with the inside circle: divide its radius into thirds, and colour the middle ring blue.

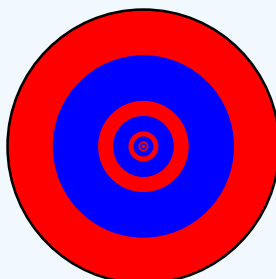


Step 1



Step 2

Continue in this way indefinitely: dividing the radius of the innermost circle into thirds, creating two rings and another circle, and colouring the middle ring blue.



What is the area of the red portion?

1. Of course we are free to define the series to be whatever we want. The hard part is defining it to be something that makes sense and doesn't lead to contradictions. We'll get to a more systematic definition shortly.
2. It is actually quite common in computer science to think of 0 as the first integer. In that context, the set of natural numbers is defined to contain 0: $\mathbb{N} = \{0, 1, 2, \dots\}$ while the notation $\mathbb{Z}^+ = \{1, 2, 3, \dots\}$ is used to denote the (strictly) positive integers. Remember that in this text, as is more standard in mathematics, we define the set of natural numbers to be the set of (strictly) positive integers.
3. This reminds the authors of the paradox of Hilbert's hotel. The hotel with an infinite number of rooms is completely full, but can always accommodate one more guest. The interested reader should use their favourite search engine to find more information on this.
4. One can find similar properties of other special series, that allow us, with some work, to cancel many terms in the partial sums. We will shortly see a good example of this. The interested reader should look up "creative telescoping" to see how this idea might be used more generally, though it is somewhat beyond this course.
5. We have included a (more) formal proof of this fact in the optional §3.7 at the end of this chapter. Proving that a repeating decimal expansion gives a rational number isn't too hard. Proving the converse — that every rational number has a repeating

decimal expansion is a little trickier, but we also do that in the same optional section.

6. Well... this sort of series does show up when you start to look at the Maclaurin polynomial of functions like $(1 - x) \log(1 - x)$. So it is not *totally* artificial. At any rate, it illustrates the basic idea of telescoping very nicely, and the idea of “creative telescoping” turns out to be extremely useful in the study of series — though it is well beyond the scope of this course.

This page titled [3.2: Series](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

3.3: Convergence Tests

It is very common to encounter series for which it is difficult, or even virtually impossible, to determine the sum exactly. Often you try to evaluate the sum approximately by truncating it, i.e. having the index run only up to some finite N , rather than infinity. But there is no point in doing so if the series diverges^{1,2}. So you like to at least know if the series converges or diverges. Furthermore you would also like to know what error is introduced when you approximate $\sum_{n=1}^{\infty} a_n$ by the “truncated series” $\sum_{n=1}^N a_n$. That's called the truncation error. There are a number of “convergence tests” to help you with this.

The Divergence Test

Our first test is very easy to apply, but it is also rarely useful. It just allows us to quickly reject some “trivially divergent” series. It is based on the observation that

- by definition, a series $\sum_{n=1}^{\infty} a_n$ converges to S when the partial sums $S_N = \sum_{n=1}^N a_n$ converge to S .
- Then, as $N \rightarrow \infty$, we have $S_N \rightarrow S$ and, because $N - 1 \rightarrow \infty$ too, we also have $S_{N-1} \rightarrow S$.
- So $a_N = S_N - S_{N-1} \rightarrow S - S = 0$.

This tells us that, if we already know that a given series $\sum a_n$ is convergent, then the n^{th} term of the series, a_n , must converge to 0 as n tends to infinity. In this form, the test is not so useful. However the contrapositive³ of the statement is a useful test for divergence.

Theorem 3.3.1 Divergence Test

If the sequence $\{a_n\}_{n=1}^{\infty}$ fails to converge to zero as $n \rightarrow \infty$, then the series $\sum_{n=1}^{\infty} a_n$ diverges.

Example 3.3.2 A simple divergence

Let $a_n = \frac{n}{n+1}$. Then

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \frac{n}{n+1} = \lim_{n \rightarrow \infty} \frac{1}{1 + \frac{1}{n}} = 1 \neq 0$$

So the series $\sum_{n=1}^{\infty} \frac{n}{n+1}$ diverges.

Warning 3.3.3

The divergence test is a “one way test”. It tells us that if $\lim_{n \rightarrow \infty} a_n$ is nonzero, or fails to exist, then the series $\sum_{n=1}^{\infty} a_n$ diverges. But it tells us *absolutely nothing* when $\lim_{n \rightarrow \infty} a_n = 0$. In particular, it is perfectly possible for a series $\sum_{n=1}^{\infty} a_n$ to *diverge* even though $\lim_{n \rightarrow \infty} a_n = 0$. An example is $\sum_{n=1}^{\infty} \frac{1}{n}$. We'll show in Example 3.3.6, below, that it diverges.

Now while convergence or divergence of series like $\sum_{n=1}^{\infty} \frac{1}{n}$ can be determined using some clever tricks — see the optional §3.3.9 —, it would be much better to have methods that are more systematic and rely less on being sneaky. Over the next subsections we will discuss several methods for testing series for convergence.

Note that while these tests will tell us whether or not a series converges, they do not (except in rare cases) tell us what the series adds up to. For example, the test we will see in the next subsection tells us quite immediately that the series

$$\sum_{n=1}^{\infty} \frac{1}{n^3}$$

converges. However it does not tell us its value⁴.

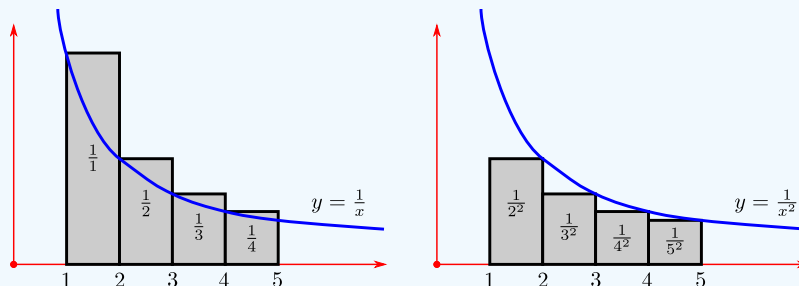
The Integral Test

In the integral test, we think of a series $\sum_{n=1}^{\infty} a_n$, that we cannot evaluate explicitly, as the area of a union of rectangles, with a_n representing the area of a rectangle of width one and height a_n . Then we compare that area with the area represented by an integral,

that we can evaluate explicitly, much as we did in Theorem 1.12.17, the comparison test for improper integrals. We'll start with a simple example, to illustrate the idea. Then we'll move on to a formulation of the test in general.

✓ Example 3.3.4 Convergence of the harmonic series

Visualise the terms of the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ as a bar graph — each term is a rectangle of height $\frac{1}{n}$ and width 1. The limit of the series is then the limiting area of this union of rectangles. Consider the sketch on the left below.



It shows that the area of the shaded columns, $\sum_{n=1}^4 \frac{1}{n}$, is bigger than the area under the curve $y = \frac{1}{x}$ with $1 \leq x \leq 5$. That is

$$\sum_{n=1}^4 \frac{1}{n} \geq \int_1^5 \frac{1}{x} dx$$

If we were to continue drawing the columns all the way out to infinity, then we would have

$$\sum_{n=1}^{\infty} \frac{1}{n} \geq \int_1^{\infty} \frac{1}{x} dx$$

We are able to compute this improper integral exactly:

$$\int_1^{\infty} \frac{1}{x} dx = \lim_{R \rightarrow \infty} [\log |x|]_1^R = +\infty$$

That is the area under the curve diverges to $+\infty$ and so the area represented by the columns must also diverge to $+\infty$.

It should be clear that the above argument can be quite easily generalised. For example the same argument holds *mutatis mutandis*⁵ for the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

Indeed we see from the sketch on the right above that

$$\sum_{n=2}^N \frac{1}{n^2} \leq \int_1^N \frac{1}{x^2} dx$$

and hence

$$\sum_{n=2}^{\infty} \frac{1}{n^2} \leq \int_1^{\infty} \frac{1}{x^2} dx$$

This last improper integral is easy to evaluate:

$$\begin{aligned} \int_2^{\infty} \frac{1}{x^2} dx &= \lim_{R \rightarrow \infty} \left[-\frac{1}{x} \right]_2^R \\ &= \lim_{R \rightarrow \infty} \left(\frac{1}{2} - \frac{1}{R} \right) = \frac{1}{2} \end{aligned}$$

Thus we know that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \sum_{n=2}^{\infty} \frac{1}{n^2} \leq \frac{3}{2}.$$

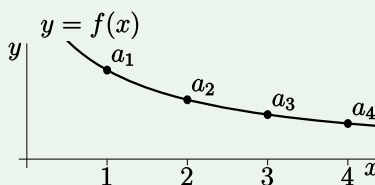
and so the series must converge.

The above arguments are formalised in the following theorem.

Theorem 3.3.5 The Integral Test

Let N_0 be any natural number. If $f(x)$ is a function which is defined and continuous for all $x \geq N_0$ and which obeys

1. $f(x) \geq 0$ for all $x \geq N_0$ and
2. $f(x)$ decreases as x increases and
3. $f(n) = a_n$ for all $n \geq N_0$.



Then

$$\sum_{n=1}^{\infty} a_n \text{ converges} \iff \int_{N_0}^{\infty} f(x) dx \text{ converges}$$

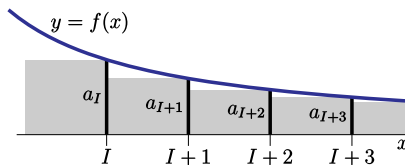
Furthermore, when the series converges, the truncation error

$$\left| \sum_{n=1}^{\infty} a_n - \sum_{n=1}^N a_n \right| \leq \int_N^{\infty} f(x) dx \quad \text{for all } N \geq N_0$$

Proof

Let I be any fixed integer with $I > N_0$. Then

- $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=I}^{\infty} a_n$ converges — removing a fixed finite number of terms from a series cannot impact whether or not it converges.
- Since $a_n \geq 0$ for all $n \geq I > N_0$, the sequence of partial sums $s_\ell = \sum_{n=I}^{\ell} a_n$ obeys $s_{\ell+1} = s_\ell + a_{\ell+1} \geq s_\ell$. That is, s_ℓ increases as ℓ increases.
- So $\{s_\ell\}$ must either converge to some finite number or increase to infinity. That is, either $\sum_{n=I}^{\infty} a_n$ converges to a finite number or it is $+\infty$.



Look at the figure above. The shaded area in the figure is $\sum_{n=I}^{\infty} a_n$

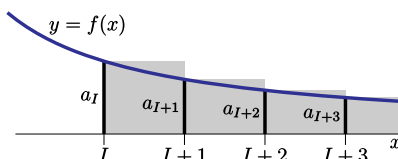
- because the first shaded rectangle has height a_I and width 1, and hence area a_I and
- the second shaded rectangle has height a_{I+1} and width 1, and hence area a_{I+1} , and so on

This shaded area is smaller than the area under the curve $y = f(x)$ for $I - 1 \leq x < \infty$. So

$$\sum_{n=I}^{\infty} a_n \leq \int_{I-1}^{\infty} f(x) dx$$

and, if the integral is finite, the sum $\sum_{n=I}^{\infty} a_n$ is finite too. Furthermore, the desired bound on the truncation error is just the special case of this inequality with $I = N + 1$:

$$\sum_{n=1}^{\infty} a_n - \sum_{n=1}^N a_n = \sum_{n=N+1}^{\infty} a_n \leq \int_N^{\infty} f(x) dx$$



For the “divergence case” look at the figure above. The (new) shaded area in the figure is again $\sum_{n=I}^{\infty} a_n$ because

- the first shaded rectangle has height a_I and width 1, and hence area a_I and
- the second shaded rectangle has height a_{I+1} and width 1, and hence area a_{I+1} , and so on

This time the shaded area is larger than the area under the curve $y = f(x)$ for $I \leq x < \infty$. So

$$\sum_{n=I}^{\infty} a_n \geq \int_I^{\infty} f(x) dx$$

and, if the integral is infinite, the sum $\sum_{n=I}^{\infty} a_n$ is infinite too.

Now that we have the integral test, it is straightforward to determine for which values of p the series ⁶

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

converges.

✓ Example 3.3.6 The p test: $\sum_{n=1}^{\infty} \frac{1}{n^p}$

Let $p > 0$. We'll now use the integral test to determine whether or not the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ (which is sometimes called the p -series) converges.

- To do so, we need a function $f(x)$ that obeys $f(n) = a_n = \frac{1}{n^p}$ for all n bigger than some N_0 . Certainly $f(x) = \frac{1}{x^p}$ obeys $f(n) = \frac{1}{n^p}$ for all $n \geq 1$. So let's pick this f and try $N_0 = 1$. (We can always increase N_0 later if we need to.)
- This function also obeys the other two conditions of Theorem 3.3.5:
 1. $f(x) > 0$ for all $x \geq N_0 = 1$ and
 2. $f(x)$ decreases as x increases because $f'(x) = -p \frac{1}{x^{p+1}} < 0$ for all $x \geq N_0 = 1$.
- So the integral test tells us that the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if and only if the integral $\int_1^{\infty} \frac{dx}{x^p}$ converges.
- We have already seen, in Example 1.12.8, that the integral $\int_1^{\infty} \frac{dx}{x^p}$ converges if and only if $p > 1$.

So we conclude that $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if and only if $p > 1$. This is sometimes called the p -test.

- In particular, the series $\sum_{n=1}^{\infty} \frac{1}{n}$, which is called the harmonic series, has $p = 1$ and so diverges. As we add more and more terms of this series together, the terms we add, namely $\frac{1}{n}$, get smaller and smaller and tend to zero, but they tend to zero so slowly that the full sum is still infinite.
- On the other hand, the series $\sum_{n=1}^{\infty} \frac{1}{n^{1.000001}}$ has $p = 1.000001 > 1$ and so converges. This time as we add more and more terms of this series together, the terms we add, namely $\frac{1}{n^{1.000001}}$, tend to zero (just) fast enough that the full sum is finite. Mind you, for this example, the convergence takes place very slowly — you have to take a huge number of terms to get a decent approximation to the full sum. If we approximate $\sum_{n=1}^{\infty} \frac{1}{n^{1.000001}}$ by the truncated series $\sum_{n=1}^N \frac{1}{n^{1.000001}}$, we make an error of at most

$$\begin{aligned} \int_N^\infty \frac{dx}{x^{1.000001}} &= \lim_{R \rightarrow \infty} \int_N^R \frac{dx}{x^{1.000001}} \\ &= \lim_{R \rightarrow \infty} -\frac{1}{0.000001} \left[\frac{1}{R^{0.000001}} - \frac{1}{N^{0.000001}} \right] \\ &= \frac{10^6}{N^{0.000001}} \end{aligned}$$

This does tend to zero as $N \rightarrow \infty$, but really slowly.

We now know that the dividing line between convergence and divergence of $\sum_{n=1}^\infty \frac{1}{n^p}$ occurs at $p = 1$. We can dig a little deeper and ask ourselves how much more quickly than $\frac{1}{n}$ the n^{th} term needs to shrink in order for the series to converge. We know that for large x , the function $\log x$ is smaller than x^a for any positive a — you can convince yourself of this with a quick application of L'Hôpital's rule. So it is not unreasonable to ask whether the series

$$\sum_{n=2}^\infty \frac{1}{n \log n}$$

converges. Notice that we sum from $n = 2$ because when $n = 1$, $n \log n = 0$. And we don't need to stop there ⁷. We can analyse the convergence of this sum with any power of $\log n$.

✓ Example 3.3.7 $\sum_{n=2}^\infty \frac{1}{n(\log n)^p}$

Let $p > 0$. We'll now use the integral test to determine whether or not the series $\sum_{n=2}^\infty \frac{1}{n(\log n)^p}$ converges.

- As in the last example, we start by choosing a function that obeys $f(n) = a_n = \frac{1}{n(\log n)^p}$ for all n bigger than some N_0 . Certainly $f(x) = \frac{1}{x(\log x)^p}$ obeys $f(n) = \frac{1}{n(\log n)^p}$ for all $n \geq 2$. So let's use that f and try $N_0 = 2$.
- Now let's check the other two conditions of Theorem 3.3.5:
 - Both x and $\log x$ are positive for all $x > 1$, so $f(x) > 0$ for all $x \geq N_0 = 2$.
 - As x increases both x and $\log x$ increase and so $x(\log x)^p$ increases and $f(x)$ decreases.
- So the integral test tells us that the series $\sum_{n=2}^\infty \frac{1}{n(\log n)^p}$ converges if and only if the integral $\int_2^\infty \frac{dx}{x(\log x)^p}$ converges.
- To test the convergence of the integral, we make the substitution $u = \log x$, $du = \frac{dx}{x}$.

$$\int_2^R \frac{dx}{x(\log x)^p} = \int_{\log 2}^{\log R} \frac{du}{u^p}$$

We already know that the integral $\int_1^\infty \frac{du}{u^p}$, and hence the integral $\int_2^R \frac{dx}{x(\log x)^p}$, converges if and only if $p > 1$.

So we conclude that $\sum_{n=2}^\infty \frac{1}{n(\log n)^p}$ converges if and only if $p > 1$.

The Comparison Test

Our next convergence test is the comparison test. It is much like the comparison test for improper integrals (see Theorem 1.12.17) and is true for much the same reasons. The rough idea is quite simple. A sum of larger terms must be bigger than a sum of smaller terms. So if we know the big sum converges, then the small sum must converge too. On the other hand, if we know the small sum diverges, then the big sum must also diverge. Formalising this idea gives the following theorem.

🔗 Theorem 3.3.8 The Comparison Test

Let N_0 be a natural number and let $K > 0$.

- If $|a_n| \leq Kc_n$ for all $n \geq N_0$ and $\sum_{n=0}^\infty c_n$ converges, then $\sum_{n=0}^\infty a_n$ converges.

2. If $a_n \geq Kd_n \geq 0$ for all $n \geq N_0$ and $\sum_{n=0}^{\infty} d_n$ diverges, then $\sum_{n=0}^{\infty} a_n$ diverges.

“Proof”

We will not prove this theorem here. We'll just observe that it is very reasonable. That's why there are quotation marks around “Proof”. For an actual proof see the optional section 3.3.10.

1. If $\sum_{n=0}^{\infty} c_n$ converges to a finite number and if the terms in $\sum_{n=0}^{\infty} a_n$ are smaller than the terms in $\sum_{n=0}^{\infty} c_n$, then it is no surprise that $\sum_{n=0}^{\infty} a_n$ converges too.
2. If $\sum_{n=0}^{\infty} d_n$ diverges (i.e. adds up to ∞) and if the terms in $\sum_{n=0}^{\infty} a_n$ are larger than the terms in $\sum_{n=0}^{\infty} d_n$, then of course $\sum_{n=0}^{\infty} a_n$ adds up to ∞ , and so diverges, too.

The comparison test for series is also used in much the same way as is the comparison test for improper integrals. Of course, one needs a good series to compare against, and often the series $\sum n^{-p}$ (from Example 3.3.6), for some $p > 0$, turns out to be just what is needed.

✓ **Example 3.3.9** $\sum_{n=1}^{\infty} \frac{1}{n^2+2n+3}$

We could determine whether or not the series $\sum_{n=1}^{\infty} \frac{1}{n^2+2n+3}$ converges by applying the integral test. But it is not worth the effort⁸. Whether or not any series converges is determined by the behaviour of the summand⁹ for very large n . So the first step in tackling such a problem is to develop some intuition about the behaviour of a_n when n is very large.

- *Step 1: Develop intuition.* In this case, when n is very large $n^2 \gg 2n \gg 3$ so that $\frac{1}{n^2+2n+3} \approx \frac{1}{n^2}$. We already know, from Example 3.3.6, that $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if and only if $p > 1$. So $\sum_{n=1}^{\infty} \frac{1}{n^2}$, which has $p = 2$, converges, and we would expect that $\sum_{n=1}^{\infty} \frac{1}{n^2+2n+3}$ converges too.
- *Step 2: Verify intuition.* We can use the comparison test to confirm that this is indeed the case. For any $n \geq 1$, $n^2 + 2n + 3 > n^2$, so that $\frac{1}{n^2+2n+3} \leq \frac{1}{n^2}$. So the comparison test, Theorem 3.3.8, with $a_n = \frac{1}{n^2+2n+3}$ and $c_n = \frac{1}{n^2}$, tells us that $\sum_{n=1}^{\infty} \frac{1}{n^2+2n+3}$ converges.

Of course the previous example was “rigged” to give an easy application of the comparison test. It is often relatively easy, using arguments like those in Example 3.3.9, to find a “simple” series $\sum_{n=1}^{\infty} b_n$ with b_n almost the same as a_n when n is large. However it is pretty rare that $a_n \leq b_n$ for all n . It is much more common that $a_n \leq Kb_n$ for some constant K . This is enough to allow application of the comparison test. Here is an example.

✓ **Example 3.3.10** $\sum_{n=1}^{\infty} \frac{n+\cos n}{n^3-1/3}$

As in the previous example, the first step is to develop some intuition about the behaviour of a_n when n is very large.

- *Step 1: Develop intuition.* When n is very large,
 - $n \gg |\cos n|$ so that the numerator $n + \cos n \approx n$ and
 - $n^3 \gg \frac{1}{3}$ so that the denominator $n^3 - \frac{1}{3} \approx n^3$.

So when n is very large

$$a_n = \frac{n + \cos n}{n^3 - \frac{1}{3}} \approx \frac{n}{n^3} = \frac{1}{n^2}$$

We already know from Example 3.3.6, with $p = 2$, that $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges, so we would expect that $\sum_{n=1}^{\infty} \frac{n+\cos n}{n^3-\frac{1}{3}}$ converges too.

- *Step 2: Verify intuition.* We can use the comparison test to confirm that this is indeed the case. To do so we need to find a constant K such that $|a_n| = \frac{|n+\cos n|}{n^3-1/3} = \frac{n+\cos n}{n^3-1/3}$ is smaller than $\frac{K}{n^2}$ for all n . A good way¹¹ to do that is to factor the dominant term (in this case n) out of the numerator and also factor the dominant term (in this case n^3) out of the denominator.

$$a_n = \frac{n + \cos n}{n^3 - \frac{1}{3}} = \frac{n}{n^3} \frac{1 + \frac{\cos n}{n}}{1 - \frac{1}{3n^3}} = \frac{1}{n^2} \frac{1 + \frac{\cos n}{n}}{1 - \frac{1}{3n^3}}$$

So now we need to find a constant K such that $\frac{1 + \frac{\cos n}{n}}{1 - \frac{1}{3n^3}}$ is smaller than K for all $n \geq 1$.

- First consider the numerator $1 + (\cos n)\frac{1}{n}$. For all $n \geq 1$

- $\frac{1}{n} \leq 1$ and
- $|\cos n| \leq 1$

So the numerator $1 + (\cos n)\frac{1}{n}$ is always smaller than $1 + (1)\frac{1}{1} = 2$.

- Next consider the denominator $1 - \frac{1}{3n^3}$.

- When $n \geq 1$, $\frac{1}{3n^3}$ lies between $\frac{1}{3}$ and 0 so that
- $1 - \frac{1}{3n^3}$ is between $\frac{2}{3}$ and 1 and consequently
- $\frac{1}{1 - \frac{1}{3n^3}}$ is between $\frac{3}{2}$ and 1.

- As the numerator $1 + (\cos n)\frac{1}{n}$ is always smaller than 2 and $\frac{1}{1 - \frac{1}{3n^3}}$ is always smaller than $\frac{3}{2}$, the fraction

$$\frac{1 + \frac{\cos n}{n}}{1 - \frac{1}{3n^3}} \leq 2 \left(\frac{3}{2} \right) = 3$$

We now know that

$$|a_n| = \frac{1}{n^2} \frac{1 + \frac{2}{n}}{1 - \frac{1}{3n^3}} \leq \frac{3}{n^2}$$

and, since we know $\sum_{n=1}^{\infty} n^{-2}$ converges, the comparison test tells us that $\sum_{n=1}^{\infty} \frac{n+\cos n}{n^3-1/3}$ converges.

The last example was actually a relatively simple application of the comparison theorem — finding a suitable constant K can be *really* tedious¹². Fortunately, there is a variant of the comparison test that completely eliminates the need to explicitly find K .

The idea behind this isn't too complicated. We have already seen that the convergence or divergence of a series depends not on its first few terms, but just on what happens when n is really large. Consequently, if we can work out how the series terms behave for really big n then we can work out if the series converges. So instead of comparing the terms of our series for all n , just compare them when n is big.

Theorem 3.3.11 Limit Comparison Theorem

Let $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ be two series with $b_n > 0$ for all n . Assume that

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = L$$

exists.

1. If $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges too.
2. If $L \neq 0$ and $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges too.

In particular, if $L \neq 0$, then $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} b_n$ converges.

Proof

(a) Because we are told that $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = L$, we know that,

- when n is large, $\frac{a_n}{b_n}$ is very close to L , so that $\left| \frac{a_n}{b_n} \right|$ is very close to $|L|$.
- In particular, there is some natural number N_0 so that $\left| \frac{a_n}{b_n} \right| \leq |L| + 1$, for all $n \geq N_0$, and hence
- $|a_n| \leq K b_n$ with $K = |L| + 1$, for all $n \geq N_0$.
- The comparison Theorem 3.3.8 now implies that $\sum_{n=1}^{\infty} a_n$ converges.

(b) Let's suppose that $L > 0$. (If $L < 0$, just replace a_n with $-a_n$.) Because we are told that $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = L$, we know that,

- when n is large, $\frac{a_n}{b_n}$ is very close to L .
- In particular, there is some natural number N so that $\frac{a_n}{b_n} \geq \frac{L}{2}$, and hence
- $a_n \geq K b_n$ with $K = \frac{L}{2} > 0$, for all $n \geq N$.
- The comparison Theorem 3.3.8 now implies that $\sum_{n=1}^{\infty} a_n$ diverges.

The next two examples illustrate how much of an improvement the above theorem is over the straight comparison test (though of course, we needed the comparison test to develop the limit comparison test).

✓ Example 3.3.12 $\sum_{n=1}^{\infty} \frac{\sqrt{n+1}}{n^2-2n+3}$

Set $a_n = \frac{\sqrt{n+1}}{n^2-2n+3}$. We first try to develop some intuition about the behaviour of a_n for large n and then we confirm that our intuition was correct.

- *Step 1: Develop intuition.* When $n \gg 1$, the numerator $\sqrt{n+1} \approx \sqrt{n}$, and the denominator $n^2 - 2n + 3 \approx n^2$ so that $a_n \approx \frac{\sqrt{n}}{n^2} = \frac{1}{n^{3/2}}$ and it looks like our series should converge by Example 3.3.6 with $p = \frac{3}{2}$.
- *Step 2: Verify intuition.* To confirm our intuition we set $b_n = \frac{1}{n^{3/2}}$ and compute the limit

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \lim_{n \rightarrow \infty} \frac{\frac{\sqrt{n+1}}{n^2-2n+3}}{\frac{1}{n^{3/2}}} = \lim_{n \rightarrow \infty} \frac{n^{3/2} \sqrt{n+1}}{n^2 - 2n + 3}$$

Again it is a good idea to factor the dominant term out of the numerator and the dominant term out of the denominator.

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \lim_{n \rightarrow \infty} \frac{n^2 \sqrt{1 + \frac{1}{n}}}{n^2 \left(1 - \frac{2}{n} + \frac{3}{n^2}\right)} = \lim_{n \rightarrow \infty} \frac{\sqrt{1 + \frac{1}{n}}}{1 - \frac{2}{n} + \frac{3}{n^2}} = 1$$

We already know that the series $\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} \frac{1}{n^{3/2}}$ converges by Example 3.3.6 with $p = \frac{3}{2}$. So our series converges by the limit comparison test, Theorem 3.3.11.

✓ Example 3.3.13 $\sum_{n=1}^{\infty} \frac{\sqrt{n+1}}{n^2-2n+3}$, again

We can also try to deal with the series of Example 3.3.12, using the comparison test directly. But that requires us to find K so that

$$\frac{\sqrt{n+1}}{n^2-2n+3} \leq \frac{K}{n^{3/2}}$$

We might do this by examining the numerator and denominator separately:

- The numerator isn't too bad since for all $n \geq 1$:

$$\begin{aligned} n+1 &\leq 2n && \text{and so} \\ \sqrt{n+1} &\leq \sqrt{2n} \end{aligned}$$

- The denominator is quite a bit more tricky, since we need a *lower* bound, rather than an upper bound, and we cannot just write $|n^2 - 2n + 3| \geq n^2$, which is false. Instead we have to make a more careful argument. In particular, we'd like to find N_0 and K' so that $n^2 - 2n + 3 \geq K' n^2$, i.e. $\frac{1}{n^2-2n+3} \leq \frac{1}{K' n^2}$ for all $n \geq N_0$. For $n \geq 4$, we have

$2n = \frac{1}{2}4n \leq \frac{1}{2}n \cdot n = \frac{1}{2}n^2$. So for $n \geq 4$,

$$n^2 - 2n + 3 \geq n^2 - \frac{1}{2}n^2 + 3 \geq \frac{1}{2}n^2$$

Putting the numerator and denominator back together we have

$$\frac{\sqrt{n+1}}{n^2 - 2n + 3} \leq \frac{\sqrt{2n}}{n^2/2} = 2\sqrt{2} \frac{1}{n^{3/2}} \quad \text{for all } n \geq 4$$

and the comparison test then tells us that our series converges. It is pretty clear that the approach of Example 3.3.12 was much more straightforward.

The Alternating Series Test

When the signs of successive terms in a series alternate between $+$ and $-$, like for example in $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$, the series is called an *alternating series*. More generally, the series

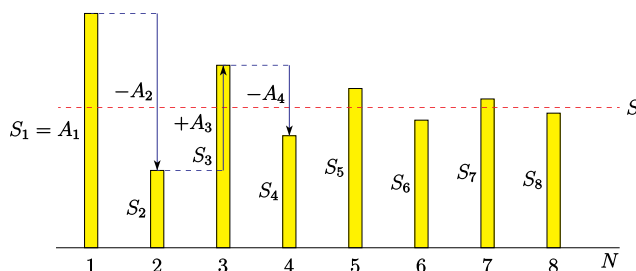
$$A_1 - A_2 + A_3 - A_4 + \dots = \sum_{n=1}^{\infty} (-1)^{n-1} A_n$$

is alternating if every $A_n \geq 0$. Often (but not always) the terms in alternating series get successively smaller. That is, then $A_1 \geq A_2 \geq A_3 \geq \dots$. In this case:

- The first partial sum is $S_1 = A_1$.
- The second partial sum, $S_2 = A_1 - A_2$, is smaller than S_1 by A_2 .
- The third partial sum, $S_3 = S_2 + A_3$, is bigger than S_2 by A_3 , but because $A_3 \leq A_2$, S_3 remains smaller than S_1 . See the figure below.
- The fourth partial sum, $S_4 = S_3 - A_4$, is smaller than S_3 by A_4 , but because $A_4 \leq A_3$, S_4 remains bigger than S_2 . Again, see the figure below.
- And so on.

So the successive partial sums oscillate, but with ever decreasing amplitude. If, in addition, A_n tends to 0 as n tends to ∞ , the amplitude of oscillation tends to zero and the sequence S_1, S_2, S_3, \dots converges to some limit S .

This is illustrated in the figure



Here is a convergence test for alternating series that exploits this structure, and that is really easy to apply.

Theorem 3.3.14 Alternating Series Test

Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of real numbers that obeys

1. $A_n \geq 0$ for all $n \geq 1$ and
2. $A_{n+1} \leq A_n$ for all $n \geq 1$ (i.e. the sequence is monotone decreasing) and
3. $\lim_{n \rightarrow \infty} A_n = 0$.

Then

$$A_1 - A_2 + A_3 - A_4 + \dots = \sum_{n=1}^{\infty} (-1)^{n-1} A_n = S$$

converges and, for each natural number N , $S - S_N$ is between 0 and (the first dropped term) $(-1)^N A_{N+1}$. Here S_N is, as previously, the N^{th} partial sum $\sum_{n=1}^N (-1)^{n-1} A_n$.

“Proof”

We shall only give part of the proof here. For the rest of the proof see the optional section 3.3.10. We shall fix any natural number N and concentrate on the last statement, which gives a bound on the truncation error (which is the error introduced when you approximate the full series by the partial sum S_N)

$$\begin{aligned} E_N = S - S_N &= \sum_{n=N+1}^{\infty} (-1)^{n-1} A_n \\ &= (-1)^N \left[A_{N+1} - A_{N+2} + A_{N+3} - A_{N+4} + \dots \right] \end{aligned}$$

This is of course another series. We're going to study the partial sums

$$S_{N,\ell} = \sum_{n=N+1}^{\ell} (-1)^{n-1} A_n = (-1)^N \sum_{m=1}^{\ell-N} (-1)^{m-1} A_{N+m}$$

for that series.

- If $\ell' > N + 1$, with $\ell' - N$ even,

$$\begin{aligned} (-1)^N S_{N,\ell'} &= \overbrace{(A_{N+1} - A_{N+2})}^{\geq 0} + \overbrace{(A_{N+3} - A_{N+4})}^{\geq 0} + \dots \\ &\quad + \overbrace{(A_{\ell'-1} - A_{\ell'})}^{\geq 0} \\ &\geq 0 \end{aligned}$$

and

$$(-1)^N S_{N,\ell'+1} = \overbrace{(-1)^N S_{N,\ell'}}^{\geq 0} + \overbrace{A_{\ell'+1}}^{\geq 0} \geq 0$$

This tells us that $(-1)^N S_{N,\ell} \geq 0$ for all $\ell > N + 1$, both even and odd.

- Similarly, if $\ell' > N + 1$, with $\ell' - N$ odd,

$$\begin{aligned} (-1)^N S_{N,\ell'} &= A_{N+1} - \overbrace{(A_{N+2} - A_{N+3})}^{\geq 0} - \overbrace{(A_{N+4} - A_{N+5})}^{\geq 0} - \dots \\ &\quad - \overbrace{(A_{\ell'-1} - A_{\ell'})}^{\geq 0} \\ &\leq A_{N+1} \\ (-1)^N S_{N,\ell'+1} &= \overbrace{(-1)^N S_{N,\ell'}}^{\leq A_{N+1}} - \overbrace{A_{\ell'+1}}^{\geq 0} \leq A_{N+1} \end{aligned}$$

This tells us that $(-1)^N S_{N,\ell} \leq A_{N+1}$ for all for all $\ell > N + 1$, both even and odd.

So we now know that $S_{N,\ell}$ lies between its first term, $(-1)^N A_{N+1}$, and 0 for all $\ell > N + 1$. While we are not going to prove it here (see the optional section 3.3.10), this implies that, since $A_{N+1} \rightarrow 0$ as $N \rightarrow \infty$, the series converges and that

$$S - S_N = \lim_{\ell \rightarrow \infty} S_{N,\ell}$$

lies between $(-1)^N A_{N+1}$ and 0.

✓ Example 3.3.15 Convergence of the alternating harmonic series

We have already seen, in Example 3.3.6, that the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges. On the other hand, the series $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$ converges by the alternating series test with $A_n = \frac{1}{n}$. Note that

1. $A_n = \frac{1}{n} \geq 0$ for all $n \geq 1$, so that $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$ really is an alternating series, and
2. $A_n = \frac{1}{n}$ decreases as n increases, and
3. $\lim_{n \rightarrow \infty} A_n = \lim_{n \rightarrow \infty} \frac{1}{n} = 0$.

so that all of the hypotheses of the alternating series test, i.e. of Theorem 3.3.14, are satisfied. We shall see, in Example 3.5.20, that

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = \log 2.$$

✓ Example 3.3.16 e

You may already know that $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. In any event, we shall prove this in Example 3.6.5, below. In particular

$$\frac{1}{e} = e^{-1} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \dots$$

is an alternating series and satisfies all of the conditions of the alternating series test, Theorem 3.3.14a:

1. The terms in the series alternate in sign.
2. The magnitude of the n^{th} term in the series decreases monotonically as n increases.
3. The n^{th} term in the series converges to zero as $n \rightarrow \infty$.

So the alternating series test guarantees that, if we approximate, for example,

$$\frac{1}{e} \approx \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \frac{1}{7!} + \frac{1}{8!} - \frac{1}{9!}$$

then the error in this approximation lies between 0 and the next term in the series, which is $\frac{1}{10!}$. That is

$$\begin{aligned} \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \frac{1}{7!} + \frac{1}{8!} - \frac{1}{9!} &\leq \frac{1}{e} \\ &\leq \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \frac{1}{7!} + \frac{1}{8!} - \frac{1}{9!} + \frac{1}{10!} \end{aligned}$$

so that

$$\begin{aligned} \frac{1}{\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \frac{1}{7!} + \frac{1}{8!} - \frac{1}{9!} + \frac{1}{10!}} &\leq e \\ &\leq \frac{1}{\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \frac{1}{7!} + \frac{1}{8!} - \frac{1}{9!}} \end{aligned}$$

which, to seven decimal places says

$$2.7182816 \leq e \leq 2.7182837$$

(To seven decimal places $e = 2.7182818$.)

The alternating series test tells us that, for any natural number N , the error that we make when we approximate $\frac{1}{e}$ by the partial sum $S_N = \sum_{n=0}^N \frac{(-1)^n}{n!}$ has magnitude no larger than $\frac{1}{(N+1)!}$. This tends to zero spectacularly quickly as N increases, simply because $(N+1)!$ increases spectacularly quickly as N increases¹³. For example $20! \approx 2.4 \times 10^{27}$.

✓ Example 3.3.17 Computing $\log \frac{11}{10}$

We will shortly see, in Example 3.5.20, that if $-1 < x \leq 1$, then

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}$$

Suppose that we have to compute $\log \frac{11}{10}$ to within an accuracy of 10^{-12} . Since $\frac{11}{10} = 1 + \frac{1}{10}$, we can get $\log \frac{11}{10}$ by evaluating $\log(1+x)$ at $x = \frac{1}{10}$, so that

$$\begin{aligned} \log \frac{11}{10} &= \log \left(1 + \frac{1}{10} \right) = \frac{1}{10} - \frac{1}{2 \times 10^2} + \frac{1}{3 \times 10^3} - \frac{1}{4 \times 10^4} + \dots \\ &= \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n \times 10^n} \end{aligned}$$

By the alternating series test, this series converges. Also by the alternating series test, approximating $\log \frac{11}{10}$ by throwing away all but the first N terms

$$\begin{aligned} \log \frac{11}{10} &\approx \frac{1}{10} - \frac{1}{2 \times 10^2} + \frac{1}{3 \times 10^3} - \frac{1}{4 \times 10^4} + \dots + (-1)^{N-1} \frac{1}{N \times 10^N} \\ &= \sum_{n=1}^N (-1)^{n-1} \frac{1}{n \times 10^n} \end{aligned}$$

introduces an error whose magnitude is no more than the magnitude of the first term that we threw away.

$$\text{error} \leq \frac{1}{(N+1) \times 10^{N+1}}$$

To achieve an error that is no more than 10^{-12} , we have to choose N so that

$$\frac{1}{(N+1) \times 10^{N+1}} \leq 10^{-12}$$

The best way to do so is simply to guess — we are not going to be able to manipulate the inequality $\frac{1}{(N+1) \times 10^{N+1}} \leq \frac{1}{10^{12}}$ into the form $N \leq \dots$, and even if we could, it would not be worth the effort. We need to choose N so that the denominator $(N+1) \times 10^{N+1}$ is at least 10^{12} . That is easy, because the denominator contains the factor 10^{N+1} which is at least 10^{12} whenever $N+1 \geq 12$, i.e. whenever $N \geq 11$. So we will achieve an error of less than 10^{-12} if we choose $N = 11$.

$$\frac{1}{(N+1) \times 10^{N+1}} \Big|_{N=11} = \frac{1}{12 \times 10^{12}} < \frac{1}{10^{12}}$$

This is not the smallest possible choice of N , but in practice that just doesn't matter — your computer is not going to care whether or not you ask it to compute a few extra terms. If you really need the smallest N that obeys $\frac{1}{(N+1) \times 10^{N+1}} \leq \frac{1}{10^{12}}$, you can next just try $N = 10$, then $N = 9$, and so on.

$$\begin{aligned} \frac{1}{(N+1) \times 10^{N+1}} \Big|_{N=11} &= \frac{1}{12 \times 10^{12}} < \frac{1}{10^{12}} \\ \frac{1}{(N+1) \times 10^{N+1}} \Big|_{N=10} &= \frac{1}{11 \times 10^{11}} < \frac{1}{10 \times 10^{11}} = \frac{1}{10^{12}} \\ \frac{1}{(N+1) \times 10^{N+1}} \Big|_{N=9} &= \frac{1}{10 \times 10^{10}} = \frac{1}{10^{11}} > \frac{1}{10^{12}} \end{aligned}$$

So in this problem, the smallest acceptable $N = 10$.

The Ratio Test

The idea behind the ratio test comes from a reexamination of the geometric series. Recall that the geometric series

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} ar^n$$

converges when $|r| < 1$ and diverges otherwise. So the convergence of this series is completely determined by the number r . This number is just the ratio of successive terms — that is $r = a_{n+1}/a_n$.

In general the ratio of successive terms of a series, $\frac{a_{n+1}}{a_n}$, is not constant, but depends on n . However, as we have noted above, the convergence of a series $\sum a_n$ is determined by the behaviour of its terms when n is large. In this way, the behaviour of this ratio when n is small tells us nothing about the convergence of the series, but the limit of the ratio as $n \rightarrow \infty$ does. This is the basis of the ratio test.

Theorem 3.3.18 Ratio Test

Let N be any positive integer and assume that $a_n \neq 0$ for all $n \geq N$.

1. If $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = L < 1$, then $\sum_{n=1}^{\infty} a_n$ converges.
2. If $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = L > 1$, or $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = +\infty$, then $\sum_{n=1}^{\infty} a_n$ diverges.

Warning 3.3.19

Beware that the ratio test provides absolutely no conclusion about the convergence or divergence of the series $\sum_{n=1}^{\infty} a_n$ if

$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = 1$. See Example 3.3.22, below.

Proof

(a) Pick any number R obeying $L < R < 1$. We are assuming that $\left| \frac{a_{n+1}}{a_n} \right|$ approaches L as $n \rightarrow \infty$. In particular there must be some natural number M so that $\left| \frac{a_{n+1}}{a_n} \right| \leq R$ for all $n \geq M$. So $|a_{n+1}| \leq R|a_n|$ for all $n \geq M$. In particular

$$\begin{aligned} |a_{M+1}| &\leq R|a_M| \\ |a_{M+2}| &\leq R|a_{M+1}| && \leq R^2|a_M| \\ |a_{M+3}| &\leq R|a_{M+2}| && \leq R^3|a_M| \\ &\vdots \\ |a_{M+\ell}| &\leq R^\ell|a_M| \end{aligned}$$

for all $\ell \geq 0$. The series $\sum_{\ell=0}^{\infty} R^\ell |a_M|$ is a geometric series with ratio R smaller than one in magnitude and so converges. Consequently, by the comparison test with a_n replaced by $A_\ell = a_{M+\ell}$ and c_n replaced by $C_\ell = R^\ell |a_M|$, the series $\sum_{\ell=1}^{\infty} a_{M+\ell} = \sum_{n=M+1}^{\infty} a_n$ converges. So the series $\sum_{n=1}^{\infty} a_n$ converges too.

(b) We are assuming that $\left| \frac{a_{n+1}}{a_n} \right|$ approaches $L > 1$ as $n \rightarrow \infty$. In particular there must be some natural number $M > N$ so that $\left| \frac{a_{n+1}}{a_n} \right| \geq 1$ for all $n \geq M$. So $|a_{n+1}| \geq |a_n|$ for all $n \geq M$. That is, $|a_n|$ increases as n increases as long as $n \geq M$. So $|a_n| \geq |a_M|$ for all $n \geq M$ and a_n cannot converge to zero as $n \rightarrow \infty$. So the series diverges by the divergence test.

Example 3.3.20 $\sum_{n=0}^{\infty} ax^{n-1}$

Fix any two nonzero real numbers a and x . We have already seen in Example 3.2.4 and Lemma 3.2.5 — we have just renamed r to x — that the geometric series $\sum_{n=0}^{\infty} ax^n$ converges when $|x| < 1$ and diverges when $|x| \geq 1$. We are now going to consider a new series, constructed by differentiating¹⁴ each term in the geometric series $\sum_{n=0}^{\infty} ax^n$. This new series is

$$\sum_{n=0}^{\infty} a_n \quad \text{with} \quad a_n = a n x^{n-1}$$

Let's apply the ratio test.

$$\left| \frac{a_{n+1}}{a_n} \right| = \left| \frac{a(n+1)x^n}{a n x^{n-1}} \right| = \frac{n+1}{n} |x| = \left(1 + \frac{1}{n}\right) |x| \rightarrow L = |x| \quad \text{as } n \rightarrow \infty$$

The ratio test now tells us that the series $\sum_{n=0}^{\infty} a n x^{n-1}$ converges if $|x| < 1$ and diverges if $|x| > 1$. It says nothing about the cases $x = \pm 1$. But in both of those cases $a_n = a n (\pm 1)^n$ does not converge to zero as $n \rightarrow \infty$ and the series diverges by the divergence test.

Notice that in the above example, we had to apply another convergence test in addition to the ratio test. This will be commonplace when we reach power series and Taylor series — the ratio test will tell us something like

| *The series converges for $|x| < R$ and diverges for $|x| > R$.*

Of course, we will still have to determine what happens when $x = +R, -R$. To determine convergence or divergence in those cases we will need to use one of the other tests we have seen.

✓ **Example 3.3.21** $\sum_{n=0}^{\infty} \frac{a}{n+1} X^{n+1}$

Once again, fix any two nonzero real numbers a and X . We again start with the geometric series $\sum_{n=0}^{\infty} a x^n$ but this time we construct a new series by integrating¹⁵ each term, $a x^n$, from $x = 0$ to $x = X$ giving $\frac{a}{n+1} X^{n+1}$. The resulting new series is

$$\sum_{n=0}^{\infty} a_n \quad \text{with} \quad a_n = \frac{a}{n+1} X^{n+1}$$

To apply the ratio test we need to compute

$$\left| \frac{a_{n+1}}{a_n} \right| = \left| \frac{\frac{a}{n+2} X^{n+2}}{\frac{a}{n+1} X^{n+1}} \right| = \frac{n+1}{n+2} |X| = \frac{1 + \frac{1}{n}}{1 + \frac{2}{n}} |X| \rightarrow L = |X| \quad \text{as } n \rightarrow \infty$$

The ratio test now tells us that the series $\sum_{n=0}^{\infty} \frac{a}{n+1} X^{n+1}$ converges if $|X| < 1$ and diverges if $|X| > 1$. It says nothing about the cases $X = \pm 1$.

If $X = 1$, the series reduces to

$$\sum_{n=0}^{\infty} \frac{a}{n+1} X^{n+1} \Big|_{X=1} = \sum_{n=0}^{\infty} \frac{a}{n+1} = a \sum_{m=1}^{\infty} \frac{1}{m} \quad \text{with } m = n+1$$

which is just a times the harmonic series, which we know diverges, by Example 3.3.6.

If $X = -1$, the series reduces to

$$\sum_{n=0}^{\infty} \frac{a}{n+1} X^{n+1} \Big|_{X=-1} = \sum_{n=0}^{\infty} (-1)^{n+1} \frac{a}{n+1}$$

which converges by the alternating series test. See Example 3.3.15.

In conclusion, the series $\sum_{n=0}^{\infty} \frac{a}{n+1} X^{n+1}$ converges if and only if $-1 \leq X < 1$.

The ratio test is often quite easy to apply, but one must always be careful when the limit of the ratio is 1. The next example illustrates this.

✓ **Example 3.3.22** $L = 1$

In this example, we are going to see three different series that all have $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = 1$. One is going to diverge and the other two are going to converge.

- The first series is the harmonic series

$$\sum_{n=1}^{\infty} a_n \quad \text{with } a_n = \frac{1}{n}$$

We have already seen, in Example 3.3.6, that this series diverges. It has

$$\left| \frac{a_{n+1}}{a_n} \right| = \left| \frac{\frac{1}{n+1}}{\frac{1}{n}} \right| = \frac{n}{n+1} = \frac{1}{1 + \frac{1}{n}} \rightarrow L = 1 \quad \text{as } n \rightarrow \infty$$

- The second series is the alternating harmonic series

$$\sum_{n=1}^{\infty} a_n \quad \text{with } a_n = (-1)^{n-1} \frac{1}{n}$$

We have already seen, in Example 3.3.15, that this series converges. But it also has

$$\left| \frac{a_{n+1}}{a_n} \right| = \left| \frac{(-1)^n \frac{1}{n+1}}{(-1)^{n-1} \frac{1}{n}} \right| = \frac{n}{n+1} = \frac{1}{1 + \frac{1}{n}} \rightarrow L = 1 \quad \text{as } n \rightarrow \infty$$

- The third series is

$$\sum_{n=1}^{\infty} a_n \quad \text{with } a_n = \frac{1}{n^2}$$

We have already seen, in Example 3.3.6 with $p = 2$, that this series converges. But it also has

$$\left| \frac{a_{n+1}}{a_n} \right| = \left| \frac{\frac{1}{(n+1)^2}}{\frac{1}{n^2}} \right| = \frac{n^2}{(n+1)^2} = \frac{1}{(1 + \frac{1}{n})^2} \rightarrow L = 1 \quad \text{as } n \rightarrow \infty$$

Let's do a somewhat artificial example that forces us to combine a few of the techniques we have seen.

✓ Example 3.3.23 $\sum_{n=1}^{\infty} \frac{(-3)^n \sqrt{n+1}}{2n+3} x^n$

Again, the convergence of this series will depend on x .

- Let us start with the ratio test — so we compute

$$\begin{aligned} \left| \frac{a_{n+1}}{a_n} \right| &= \left| \frac{(-3)^{n+1} \sqrt{n+2} (2n+3) x^{n+1}}{(-3)^n \sqrt{n+1} (2n+5) x^n} \right| \\ &= |-3| \cdot \frac{\sqrt{n+2}}{\sqrt{n+1}} \cdot \frac{2n+3}{2n+5} \cdot |x| \end{aligned}$$

So in the limit as $n \rightarrow \infty$ we are left with

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = 3|x|$$

- The ratio test then tells us that if $3|x| > 1$ the series diverges, while when $3|x| < 1$ the series converges.
- This leaves us with the cases $x = +\frac{1}{3}$ and $-\frac{1}{3}$.
- Setting $x = \frac{1}{3}$ gives the series

$$\sum_{n=1}^{\infty} \frac{(-1)^n \sqrt{n+1}}{2n+3}$$

The fact that the terms alternate here suggests that we use the alternating series test. That will show that this series converges provided $\frac{\sqrt{n+1}}{2n+3}$ decreases as n increases. So we define the function

$$f(t) = \frac{\sqrt{t+1}}{2t+3}$$

(which is constructed by replacing the n in $\frac{\sqrt{n+1}}{2n+3}$ with t) and verify that $f(t)$ is a decreasing function of t . To prove that, it suffices to show its derivative is negative when $t \geq 1$:

$$\begin{aligned} f'(t) &= \frac{(2t+3) \cdot \frac{1}{2} \cdot (t+1)^{-1/2} - 2\sqrt{t+1}}{(2t+3)^2} \\ &= \frac{(2t+3) - 4(t+1)}{2\sqrt{t+1}(2t+3)^2} \\ &= \frac{-2t-1}{2\sqrt{t+1}(2t+3)^2} \end{aligned}$$

When $t \geq 1$ this is negative and so $f(t)$ is a decreasing function. Thus we can apply the alternating series test to show that the series converges when $x = \frac{1}{3}$.

- When $x = -\frac{1}{3}$ the series becomes

$$\sum_{n=1}^{\infty} \frac{\sqrt{n+1}}{2n+3}.$$

Notice that when n is large, the summand is approximately $\frac{\sqrt{n}}{2n}$ which suggests that the series will diverge by comparison with $\sum n^{-1/2}$. To formalise this, we can use the limit comparison theorem:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\sqrt{n+1}}{2n+3} \frac{1}{n^{-1/2}} &= \lim_{n \rightarrow \infty} \frac{\sqrt{n} \cdot \sqrt{1+1/n}}{n(2+3/n)} \cdot n^{1/2} \\ &= \lim_{n \rightarrow \infty} \frac{n \cdot \sqrt{1+1/n}}{n(2+3/n)} \\ &= \frac{1}{2} \end{aligned}$$

So since this ratio has a finite limit and the series $\sum n^{-1/2}$ diverges, we know that our series also diverges.

So in summary the series converges when $-\frac{1}{3} < x \leq \frac{1}{3}$ and diverges otherwise.

Convergence Test List

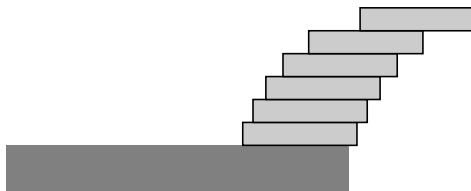
We now have half a dozen convergence tests:

- **Divergence Test**
 - works well when the n^{th} term in the series *fails* to converge to zero as n tends to infinity
- **Alternating Series Test**
 - works well when successive terms in the series alternate in sign
 - don't forget to check that successive terms decrease in magnitude and tend to zero as n tends to infinity
- **Integral Test**
 - works well when, if you substitute x for n in the n^{th} term you get a function, $f(x)$, that you can integrate
 - don't forget to check that $f(x) \geq 0$ and that $f(x)$ decreases as x increases
- **Ratio Test**
 - works well when $\frac{a_{n+1}}{a_n}$ simplifies enough that you can easily compute $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = L$
 - this often happens when a_n contains powers, like 7^n , or factorials, like $n!$
 - don't forget that $L = 1$ tells you nothing about the convergence/divergence of the series
- **Comparison Test and Limit Comparison Test**
 - works well when, for very large n , the n^{th} term a_n is approximately the same as a simpler term b_n (see Example 3.3.10) and it is easy to determine whether or not $\sum_{n=1}^{\infty} b_n$ converges

- don't forget to check that $b_n \geq 0$
- usually the Limit Comparison Test is easier to apply than the Comparison Test

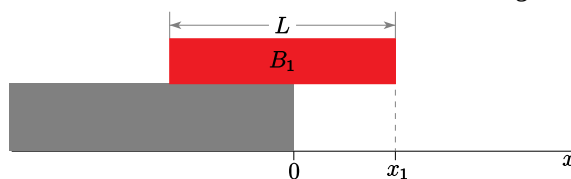
Optional — The Leaning Tower of Books

Imagine that you are about to stack a bunch of identical books on a table. But you don't want to just stack them exactly vertically. You want to build a “leaning tower of books” that overhangs the edge of the table as much as possible.



How big an overhang can you get? The answer to that question, which we'll now derive, uses a series!

- Let's start by just putting book #1 on the table. It's the red book labelled “ B_1 ” in the figure below.



Use a horizontal x -axis with $x = 0$ corresponding to the right hand edge of the table. Imagine that we have placed book #1 so that its right hand edge overhangs the end of the table by a distance x_1 .

- In order for the book to not topple off of the table, we need its center of mass to lie above the table. That is, we need the x -coordinate of the center mass of B_1 , which we shall denote $\bar{X}(B_1)$, to obey

$$\bar{X}(B_1) \leq 0$$

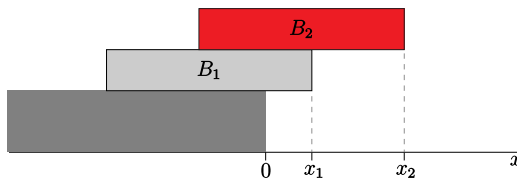
Assuming that our books have uniform density and are of length L , $\bar{X}(B_1)$ will be exactly half way between the right hand end of the book, which is at $x = x_1$, and the left hand end of the book, which is at $x = x_1 - L$. So

$$\bar{X}(B_1) = \frac{1}{2}x_1 + \frac{1}{2}(x_1 - L) = x_1 - \frac{L}{2}$$

Thus book #1 does not topple off of the table provided

$$x_1 \leq \frac{L}{2}$$

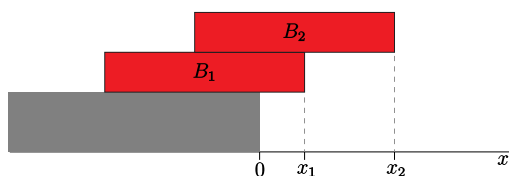
- Now let's put books #1 and #2 on the table, with the right hand edge of book #1 at $x = x_1$ and the right hand edge of book #2 at $x = x_2$, as in the figure below.



- In order for book #2 to not topple off of book #1, we need the center of mass of book #2 to lie above book #1. That is, we need the x -coordinate of the center mass of B_2 , which is $\bar{X}(B_2) = x_2 - \frac{L}{2}$, to obey

$$\bar{X}(B_2) \leq x_1 \iff x_2 - \frac{L}{2} \leq x_1 \iff x_2 \leq x_1 + \frac{L}{2}$$

- Assuming that book #2 does not topple off of book #1, we still need to arrange that the pair of books does not topple off of the table. Think of the pair of books as the combined red object in the figure



In order for the combined red object to not topple off of the table, we need the center of mass of the combined red object to lie above the table. That is, we need the x -coordinate of the center mass of the combined red object, which we shall denote $\bar{X}(B_1 \cup B_2)$, to obey

$$\bar{X}(B_1 \cup B_2) \leq 0$$

The center of mass of the combined red object is the weighted average¹⁶ of the centers of mass of B_1 and B_2 . As B_1 and B_2 have the same weight,

$$\begin{aligned} \bar{X}(B_1 \cup B_2) &= \frac{1}{2}\bar{X}(B_1) + \frac{1}{2}\bar{X}(B_2) = \frac{1}{2}\left(x_1 - \frac{L}{2}\right) + \frac{1}{2}\left(x_2 - \frac{L}{2}\right) \\ &= \frac{1}{2}(x_1 + x_2) - \frac{L}{2} \end{aligned}$$

and the combined red object does not topple off of the table if

$$\bar{X}(B_1 \cup B_2) = \frac{1}{2}(x_1 + x_2) - \frac{L}{2} \leq 0 \iff x_1 + x_2 \leq L$$

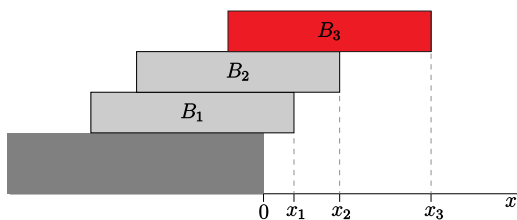
In conclusion, our two-book tower survives if

$$x_2 \leq x_1 + \frac{L}{2} \quad \text{and} \quad x_1 + x_2 \leq L$$

In particular we may choose x_1 and x_2 to satisfy $x_2 = x_1 + \frac{L}{2}$ and $x_1 + x_2 = L$. Then, substituting $x_2 = x_1 + \frac{L}{2}$ into $x_1 + x_2 = L$ gives

$$x_1 + \left(x_1 + \frac{L}{2}\right) = L \iff 2x_1 = \frac{L}{2} \iff x_1 = \frac{L}{2}\left(\frac{1}{2}\right), \quad x_2 = \frac{L}{2}\left(1 + \frac{1}{2}\right)$$

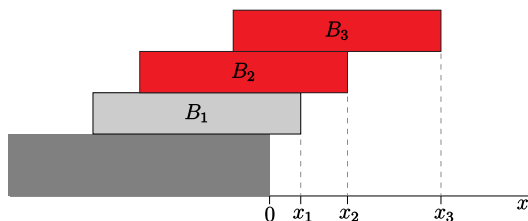
- Before considering the general “ n -book tower”, let’s now put books #1, #2 and #3 on the table, with the right hand edge of book #1 at $x = x_1$, the right hand edge of book #2 at $x = x_2$, and the right hand edge of book #3 at $x = x_3$, as in the figure below.



- In order for book #3 to not topple off of book #2, we need the center of mass of book #3 to lie above book #2. That is, we need the x -coordinate of the center mass of B_3 , which is $\bar{X}(B_3) = x_3 - \frac{L}{2}$, to obey

$$\bar{X}(B_3) \leq x_2 \iff x_3 - \frac{L}{2} \leq x_2 \iff x_3 \leq x_2 + \frac{L}{2}$$

- Assuming that book #3 does not topple off of book #2, we still need to arrange that the pair of books, book #2 plus book #3 (the red object in the figure below), does not topple off of book #1.



In order for this combined red object to not topple off of book #1, we need the x -coordinate of its center mass, which we denote $\bar{X}(B_2 \cup B_3)$, to obey

$$\bar{X}(B_2 \cup B_3) \leq x_1$$

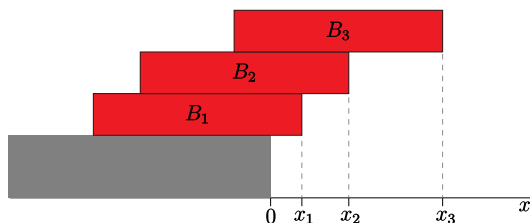
The center of mass of the combined red object is the weighted average of the center of masses of B_2 and B_3 . As B_2 and B_3 have the same weight,

$$\begin{aligned} \bar{X}(B_2 \cup B_3) &= \frac{1}{2}\bar{X}(B_2) + \frac{1}{2}\bar{X}(B_3) = \frac{1}{2}\left(x_2 - \frac{L}{2}\right) + \frac{1}{2}\left(x_3 - \frac{L}{2}\right) \\ &= \frac{1}{2}(x_2 + x_3) - \frac{L}{2} \end{aligned}$$

and the combined red object does not topple off of book #1 if

$$\frac{1}{2}(x_2 + x_3) - \frac{L}{2} \leq x_1 \iff x_2 + x_3 \leq 2x_1 + L$$

- Assuming that book #3 does not topple off of book #2, and also that the combined book #2 plus book #3 does not topple off of book #1, we still need to arrange that the whole tower of books, book #1 plus book #2 plus book #3 (the red object in the figure below), does not topple off of the table.



In order for this combined red object to not topple off of the table, we need the x -coordinate of its center mass, which we denote $\bar{X}(B_1 \cup B_2 \cup B_3)$, to obey

$$\bar{X}(B_1 \cup B_2 \cup B_3) \leq 0$$

The center of mass of the combined red object is the weighted average of the center of masses of B_1 and B_2 and B_3 . As they all have the same weight,

$$\begin{aligned} \bar{X}(B_1 \cup B_2 \cup B_3) &= \frac{1}{3}\bar{X}(B_1) + \frac{1}{3}\bar{X}(B_2) + \frac{1}{3}\bar{X}(B_3) \\ &= \frac{1}{3}\left(x_1 - \frac{L}{2}\right) + \frac{1}{3}\left(x_2 - \frac{L}{2}\right) + \frac{1}{3}\left(x_3 - \frac{L}{2}\right) \\ &= \frac{1}{3}(x_1 + x_2 + x_3) - \frac{L}{2} \end{aligned}$$

and the combined red object does not topple off of the table if

$$\frac{1}{3}(x_1 + x_2 + x_3) - \frac{L}{2} \leq 0 \iff x_1 + x_2 + x_3 \leq \frac{3L}{2}$$

In conclusion, our three-book tower survives if

$$x_3 \leq x_2 + \frac{L}{2} \quad \text{and} \quad x_2 + x_3 \leq 2x_1 + L \quad \text{and} \quad x_1 + x_2 + x_3 \leq \frac{3L}{2}$$

In particular, we may choose x_1 , x_2 and x_3 to satisfy

$$\begin{aligned} x_1 + x_2 + x_3 &= \frac{3L}{2} \quad \text{and} \\ x_2 + x_3 &= 2x_1 + L \quad \text{and} \\ x_3 &= \frac{L}{2} + x_2 \end{aligned}$$

Substituting the second equation into the first gives

$$3x_1 + L = \frac{3L}{2} \implies x_1 = \frac{L}{2} \left(\frac{1}{3} \right)$$

Next substituting the third equation into the second, and then using the formula above for x_1 , gives

$$2x_2 + \frac{L}{2} = 2x_1 + L = \frac{L}{3} + L \implies x_2 = \frac{L}{2} \left(\frac{1}{2} + \frac{1}{3} \right)$$

and finally

$$x_3 = \frac{L}{2} + x_2 = \frac{L}{2} \left(1 + \frac{1}{2} + \frac{1}{3} \right)$$

- We are finally ready for the general “ n -book tower”. Stack n books on the table, with book B_1 on the bottom and book B_n at the top, and with the right hand edge of book $\#j$ at $x = x_j$. The same center of mass considerations as above show that the tower survives if

$$\begin{array}{rcl} \bar{X}(B_n) \leq x_{n-1} & & x_n - \frac{L}{2} \leq x_{n-1} \\ \bar{X}(B_{n-1} \cup B_n) \leq x_{n-2} & & \frac{1}{2}(x_{n-1} + x_n) - \frac{L}{2} \leq x_{n-2} \\ & \vdots & \vdots \\ \bar{X}(B_3 \cup \dots \cup B_n) \leq x_2 & & \frac{1}{n-2}(x_3 + \dots + x_n) - \frac{L}{2} \leq x_2 \\ \bar{X}(B_2 \cup B_3 \cup \dots \cup B_n) \leq x_1 & & \frac{1}{n-1}(x_2 + x_3 + \dots + x_n) - \frac{L}{2} \leq x_1 \\ \bar{X}(B_1 \cup B_2 \cup B_3 \cup \dots \cup B_n) \leq 0 & & \frac{1}{n}(x_1 + x_2 + x_3 + \dots + x_n) - \frac{L}{2} \leq 0 \end{array}$$

In particular, we may choose the x_j 's to obey

$$\begin{array}{rcl} \frac{1}{n}(x_1 + x_2 + x_3 + \dots + x_n) & = & \frac{L}{2} \\ \frac{1}{n-1}(x_2 + x_3 + \dots + x_n) & = & \frac{L}{2} + x_1 \\ \frac{1}{n-2}(x_3 + \dots + x_n) & = & \frac{L}{2} + x_2 \\ & \vdots & \vdots \\ \frac{1}{2}(x_{n-1} + x_n) & = & \frac{L}{2} + x_{n-2} \\ x_n & = & \frac{L}{2} + x_{n-1} \end{array}$$

Substituting $x_2 + x_3 + \dots + x_n = (n-1)x_1 + \frac{L}{2}(n-1)$ from the second equation into the first equation gives

$$\begin{aligned} \frac{1}{n} \left\{ \overbrace{x_1 + (n-1)x_1}^{nx_1} + \frac{L}{2}(n-1) \right\} &= \frac{L}{2} \implies x_1 + \frac{L}{2} \left(1 - \frac{1}{n} \right) = \frac{L}{2} \left(\frac{1}{2} \right) \\ &\implies x_1 = \frac{L}{2} \left(\frac{1}{n} \right) \end{aligned}$$

Substituting $x_3 + \dots + x_n = (n-2)x_2 + \frac{L}{2}(n-2)$ from the third equation into the second equation gives

$$\begin{aligned} \frac{1}{n-1} \left\{ \overbrace{x_2 + (n-2)x_2}^{(n-1)x_2} + \frac{L}{2} \overbrace{(n-2)}^{(n-1)-1} \right\} &= \frac{L}{2} + x_1 = \frac{L}{2} \left(1 + \frac{1}{n} \right) \\ \implies x_2 + \frac{L}{2} \left(1 - \frac{1}{n-1} \right) &= \frac{L}{2} \left(1 + \frac{1}{n} \right) \\ \implies x_2 = \frac{L}{2} \left(\frac{1}{n-1} + \frac{1}{n} \right) \end{aligned}$$

Just keep going. We end up with

$$\begin{aligned}
 x_1 &= \frac{L}{2} \left(\frac{1}{n} \right) \\
 x_2 &= \frac{L}{2} \left(\frac{1}{n-1} + \frac{1}{n} \right) \\
 x_3 &= \frac{L}{2} \left(\frac{1}{n-2} + \frac{1}{n-1} + \frac{1}{n} \right) \\
 &\vdots \\
 x_{n-2} &= \frac{L}{2} \left(\frac{1}{3} + \cdots + \frac{1}{n} \right) \\
 x_{n-1} &= \frac{L}{2} \left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \right) \\
 x_n &= \frac{L}{2} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \right)
 \end{aligned}$$

Our overhang is $x_n = \frac{L}{2} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \right)$. This is $\frac{L}{2}$ times the n^{th} partial sum of the harmonic series $\sum_{m=1}^{\infty} \frac{1}{m}$. As we saw in Example 3.3.6 (the p test), the harmonic series diverges. So, as n goes to infinity $1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$ also goes to infinity. We may make the overhang as large¹⁷ as we like!

Optional — The Root Test

There is another test that is very similar in spirit to the ratio test. It also comes from a reexamination of the geometric series

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} ar^n$$

The ratio test was based on the observation that r , which largely determines whether or not the series converges, could be found by computing the ratio $r = a_{n+1}/a_n$. The root test is based on the observation that $|r|$ can also be determined by looking that the n^{th} root of the n^{th} term with n very large:

$$\lim_{n \rightarrow \infty} \sqrt[n]{|ar^n|} = |r| \lim_{n \rightarrow \infty} \sqrt[n]{|a|} = |r| \quad \text{if } a \neq 0$$

Of course, in general, the n^{th} term is not exactly ar^n . However, if for very large n , the n^{th} term is approximately proportional to r^n , with $|r|$ given by the above limit, we would expect the series to converge when $|r| < 1$ and diverge when $|r| > 1$. That is indeed the case.

Theorem 3.3.24 Root Test

Assume that

$$L = \lim_{n \rightarrow \infty} \sqrt[n]{|a_n|}$$

exists or is $+\infty$.

1. If $L < 1$, then $\sum_{n=1}^{\infty} a_n$ converges.
2. If $L > 1$, or $L = +\infty$, then $\sum_{n=1}^{\infty} a_n$ diverges.

Warning 3.3.25

Beware that the root test provides absolutely no conclusion about the convergence or divergence of the series $\sum_{n=1}^{\infty} a_n$ if

$$\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = 1.$$

Proof

(a) Pick any number R obeying $L < R < 1$. We are assuming that $\sqrt[n]{|a_n|}$ approaches L as $n \rightarrow \infty$. In particular there must be some natural number M so that $\sqrt[n]{|a_n|} \leq R$ for all $n \geq M$. So $|a_n| \leq R^n$ for all $n \geq M$ and the series $\sum_{n=1}^{\infty} a_n$ converges by comparison to the geometric series $\sum_{n=1}^{\infty} R^n$

(b) We are assuming that $\sqrt[n]{|a_n|}$ approaches $L > 1$ (or grows unboundedly) as $n \rightarrow \infty$. In particular there must be some natural number M so that $\sqrt[n]{|a_n|} \geq 1$ for all $n \geq M$. So $|a_n| \geq 1$ for all $n \geq M$ and the series diverges by the divergence test.

✓ Example 3.3.26 $\sum_{n=1}^{\infty} \frac{(-3)^n \sqrt{n+1}}{2n+3} x^n$

We have already used the ratio test, in Example 3.3.23, to show that this series converges when $|x| < \frac{1}{3}$ and diverges when $|x| > \frac{1}{3}$. We'll now use the root test to draw the same conclusions.

- Write $a_n = \frac{(-3)^n \sqrt{n+1}}{2n+3} x^n$.
- We compute

$$\begin{aligned} \sqrt[n]{|a_n|} &= \sqrt[n]{\left| \frac{(-3)^n \sqrt{n+1}}{2n+3} x^n \right|} \\ &= 3|x| (n+1)^{\frac{1}{2n}} (2n+3)^{-\frac{1}{n}} \end{aligned}$$

- We'll now show that the limit of $(n+1)^{\frac{1}{2n}}$ as $n \rightarrow \infty$ is exactly 1. To do, so we first compute the limit of the logarithm.

$$\begin{aligned} \lim_{n \rightarrow \infty} \log(n+1)^{\frac{1}{2n}} &= \lim_{n \rightarrow \infty} \frac{\log(n+1)}{2n} && \text{now apply Theorem 3.1.6} \\ &= \lim_{t \rightarrow \infty} \frac{\log(t+1)}{2t} \\ &= \lim_{t \rightarrow \infty} \frac{\frac{1}{t+1}}{2} && \text{by l'Hôpital} \\ &= 0 \end{aligned}$$

So

$$\lim_{n \rightarrow \infty} (n+1)^{\frac{1}{2n}} = \lim_{n \rightarrow \infty} \exp\{\log(n+1)^{\frac{1}{2n}}\} = e^0 = 1$$

An essentially identical computation also gives that $\lim_{n \rightarrow \infty} (2n+3)^{-\frac{1}{n}} = e^0 = 1$.

- So

$$\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = 3|x|$$

and the root test also tells us that if $3|x| > 1$ the series diverges, while when $3|x| < 1$ the series converges.

We have done the last example once, in Example 3.3.23, using the ratio test and once, in Example 3.3.26, using the root test. It was clearly much easier to use the ratio test. Here is an example that is most easily handled by the root test.

✓ Example 3.3.27 $\sum_{n=1}^{\infty} \left(\frac{n}{n+1}\right)^{n^2}$

Write $a_n = \left(\frac{n}{n+1}\right)^{n^2}$. Then

$$\sqrt[n]{|a_n|} = \sqrt[n]{\left(\frac{n}{n+1}\right)^{n^2}} = \left(\frac{n}{n+1}\right)^n = \left(1 + \frac{1}{n}\right)^{-n}$$

Now we take the limit,

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^{-n} &= \lim_{X \rightarrow \infty} \left(1 + \frac{1}{X}\right)^{-X} && \text{by Theorem 3.1.6} \\ &= \lim_{x \rightarrow 0} (1+x)^{-1/x} && \text{where } x = \frac{1}{X} \\ &= e^{-1} \end{aligned}$$

by Example 3.7.20 in the CLP-1 text with $a = -1$. As the limit is strictly smaller than 1, the series $\sum_{n=1}^{\infty} \left(\frac{n}{n+1}\right)^{n^2}$ converges.

To draw the same conclusion using the ratio test, one would have to show that the limit of

$$\frac{a_{n+1}}{a_n} = \left(\frac{n+1}{n+2}\right)^{(n+1)^2} \left(\frac{n+1}{n}\right)^{n^2}$$

as $n \rightarrow \infty$ is strictly smaller than 1. It's clearly better to stick with the root test.

Optional — Harmonic and Basel Series

The Harmonic Series

The series

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

that appeared in Warning 3.3.3, is called the Harmonic series¹⁸, and its partial sums

$$H_N = \sum_{n=1}^N \frac{1}{n}$$

are called the Harmonic numbers. Though these numbers have been studied at least as far back as Pythagoras, the divergence of the series was first proved in around 1350 by Nicholas Oresme (1320-5 – 1382), though the proof was lost for many years and rediscovered by Mengoli (1626–1686) and the Bernoulli brothers (Johann 1667–1748 and Jacob 1655–1705).

Oresme's proof is beautiful and all the more remarkable that it was produced more than 300 years before calculus was developed by Newton and Leibnitz. It starts by grouping the terms of the harmonic series carefully:

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{1}{n} &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \dots \\ &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \left(\frac{1}{9} + \frac{1}{10} + \dots + \frac{1}{15} + \frac{1}{16}\right) + \dots \\ &> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \left(\frac{1}{16} + \frac{1}{16} + \dots + \frac{1}{16} + \frac{1}{16}\right) + \dots \\ &= 1 + \frac{1}{2} + \left(\frac{2}{4}\right) + \left(\frac{4}{8}\right) + \left(\frac{8}{16}\right) + \dots \end{aligned}$$

So one can see that this is $1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots$ and so must diverge¹⁹.

There are many variations on Oresme's proof — for example, using groups of two or three. A rather different proof relies on the inequality

$$e^x > 1 + x \quad \text{for } x > 0$$

which follows immediately from the Taylor series for e^x given in Theorem 3.6.7. From this we can bound the exponential of the Harmonic numbers:

$$\begin{aligned}
 e^{H_n} &= e^{1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{n}} \\
 &= e^1 \cdot e^{1/2} \cdot e^{1/3} \cdot e^{1/4} \dots e^{1/n} \\
 &> (1+1) \cdot (1+1/2) \cdot (1+1/3) \cdot (1+1/4) \dots (1+1/n) \\
 &= \frac{2}{1} \cdot \frac{3}{2} \cdot \frac{4}{3} \cdot \frac{5}{4} \dots \frac{n+1}{n} \\
 &= n+1
 \end{aligned}$$

Since e^{H_n} grows unboundedly with n , the harmonic series diverges.

The Basel Problem

The problem of determining the exact value of the sum of the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

is called the Basel problem. The problem is named after the home town of Leonhard Euler, who solved it. One can use telescoping series to show that this series must converge. Notice that

$$\frac{1}{n^2} < \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}$$

Hence we can bound the partial sum:

$$\begin{aligned}
 S_k &= \sum_{n=1}^k \frac{1}{n^2} < 1 + \sum_{n=2}^k \frac{1}{n(n-1)} && \text{avoid dividing by 0} \\
 &= 1 + \sum_{n=2}^k \left(\frac{1}{n-1} - \frac{1}{n} \right) && \text{which telescopes to} \\
 &= 1 + 1 - \frac{1}{k}
 \end{aligned}$$

Thus, as k increases, the partial sum S_k increases (the series is a sum of positive terms), but is always smaller than 2. So the sequence of partial sums converges.

Mengoli posed the problem of evaluating the series exactly in 1644 and it was solved — not entirely rigorously — by Euler in 1734. A rigorous proof had to wait another 7 years. Euler used some extremely cunning observations and manipulations of the sine function to show that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

He used the Maclaurin series

$$\sin x = 1 - \frac{x^3}{6} + \frac{x^5}{24} - \dots$$

and a product formula for sine

$$\begin{aligned}
 \sin x &= x \cdot \left(1 - \frac{x}{\pi}\right) \cdot \left(1 + \frac{x}{\pi}\right) \cdot \left(1 - \frac{x}{2\pi}\right) \cdot \left(1 + \frac{x}{2\pi}\right) \cdot \left(1 - \frac{x}{3\pi}\right) \cdot \left(1 + \frac{x}{3\pi}\right) \dots \\
 &= x \cdot \left(1 - \frac{x^2}{\pi^2}\right) \cdot \left(1 - \frac{x^2}{4\pi^2}\right) \cdot \left(1 - \frac{x^2}{9\pi^2}\right) \dots
 \end{aligned} \tag{*}$$

Extracting the coefficient of x^3 from both expansions gives the desired result. The proof of the product formula is well beyond the scope of this course. But notice that at least the values of x which make the left hand side of (*) zero, namely $x = n\pi$ with n integer, are exactly the same as the values of x which make the right hand side of (*) zero²⁰.

This approach can also be used to compute $\sum_{n=1}^{\infty} n^{-2p}$ for $p = 1, 2, 3, \dots$ and show that they are rational multiples²¹ of π^{2p} . The corresponding series of odd powers are significantly nastier and getting closed form expressions for them remains a famous open problem.

Optional — Some Proofs

In this optional section we provide proofs of two convergence tests. We shall repeatedly use the fact that any sequence a_1, a_2, a_3, \dots , of real numbers which is increasing (i.e. $a_{n+1} \geq a_n$ for all n) and bounded (i.e. there is a constant M such that $a_n \leq M$ for all n) converges. We shall not prove this fact²².

We start with the comparison test, and then move on to the alternating series test.

Theorem 3.3.28 The Comparison Test (stated again)

Let N_0 be a natural number and let $K > 0$.

1. If $|a_n| \leq Kc_n$ for all $n \geq N_0$ and $\sum_{n=0}^{\infty} c_n$ converges, then $\sum_{n=0}^{\infty} a_n$ converges.
2. If $a_n \geq Kd_n \geq 0$ for all $n \geq N_0$ and $\sum_{n=0}^{\infty} d_n$ diverges, then $\sum_{n=0}^{\infty} a_n$ diverges.

Proof

(a) By hypothesis $\sum_{n=0}^{\infty} c_n$ converges. So it suffices to prove that $\sum_{n=0}^{\infty} [Kc_n - a_n]$ converges, because then, by our Arithmetic of series Theorem 3.2.9,

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} Kc_n - \sum_{n=0}^{\infty} [Kc_n - a_n]$$

will converge too. But for all $n \geq N_0$, $Kc_n - a_n \geq 0$ so that, for all $N \geq N_0$, the partial sums

$$S_N = \sum_{n=0}^N [Kc_n - a_n]$$

increase with N , but never gets bigger than the finite number $\sum_{n=0}^{N_0} [Kc_n - a_n] + K \sum_{n=N_0+1}^{\infty} c_n$. So the partial sums S_N converge as $N \rightarrow \infty$.

(b) For all $N > N_0$, the partial sum

$$S_N = \sum_{n=0}^N a_n \geq \sum_{n=0}^{N_0} a_n + K \sum_{n=N_0+1}^N d_n$$

By hypothesis, $\sum_{n=N_0+1}^N d_n$, and hence S_N , grows without bound as $N \rightarrow \infty$. So $S_N \rightarrow \infty$ as $N \rightarrow \infty$.

Theorem 3.3.29 Alternating Series Test (stated again)

Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers that obeys

1. $a_n \geq 0$ for all $n \geq 1$ and
2. $a_{n+1} \leq a_n$ for all $n \geq 1$ (i.e. the sequence is monotone decreasing) and
3. $\lim_{n \rightarrow \infty} a_n = 0$.

Then

$$a_1 - a_2 + a_3 - a_4 + \dots = \sum_{n=1}^{\infty} (-1)^{n-1} a_n = S$$

converges and, for each natural number N , $S - S_N$ is between 0 and (the first dropped term) $(-1)^N a_{N+1}$. Here S_N is, as previously, the N^{th} partial sum $\sum_{n=1}^N (-1)^{n-1} a_n$.

Proof

Let $2n$ be an even natural number. Then the $2n^{\text{th}}$ partial sum obeys

$$\begin{aligned} S_{2n} &= \overbrace{(a_1 - a_2)}^{\geq 0} + \overbrace{(a_3 - a_4)}^{\geq 0} + \cdots + \overbrace{(a_{2n-1} - a_{2n})}^{\geq 0} \\ &\leq \overbrace{(a_1 - a_2)}^{\geq 0} + \overbrace{(a_3 - a_4)}^{\geq 0} + \cdots + \overbrace{(a_{2n-1} - a_{2n})}^{\geq 0} + \overbrace{(a_{2n+1} - a_{2n+2})}^{\geq 0} \\ &= S_{2(n+1)} \end{aligned}$$

and

$$\begin{aligned} S_{2n} &= a_1 - \overbrace{(a_2 - a_3)}^{\geq 0} - \overbrace{(a_4 - a_5)}^{\geq 0} - \cdots - \overbrace{(a_{2n-2} - a_{2n-1})}^{\geq 0} - \overbrace{a_{2n}}^{\geq 0} \\ &\leq a_1 \end{aligned}$$

So the sequence S_2, S_4, S_6, \dots of even partial sums is a bounded, increasing sequence and hence converges to some real number S . Since $S_{2n+1} = S_{2n} + a_{2n+1}$ and a_{2n+1} converges zero as $n \rightarrow \infty$, the odd partial sums S_{2n+1} also converge to S . That $S - S_N$ is between 0 and (the first dropped term) $(-1)^N a_{N+1}$ was already proved in §3.3.4.

Exercises

Stage 1

? 1

Select the series below that diverge by the divergence test.

- (A) $\sum_{n=1}^{\infty} \frac{1}{n}$
- (B) $\sum_{n=1}^{\infty} \frac{n^2}{n+1}$
- (C) $\sum_{n=1}^{\infty} \sin n$
- (D) $\sum_{n=1}^{\infty} \sin(\pi n)$

? 2

Select the series below whose terms satisfy the conditions to apply the integral test.

- (A) $\sum_{n=1}^{\infty} \frac{1}{n}$
- (B) $\sum_{n=1}^{\infty} \frac{n^2}{n+1}$
- (C) $\sum_{n=1}^{\infty} \sin n$
- (D) $\sum_{n=1}^{\infty} \frac{\sin n + 1}{n^2}$

? 3

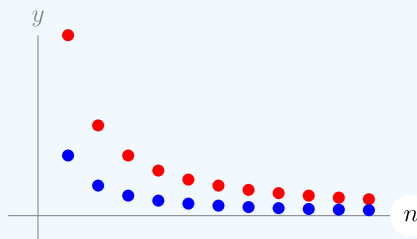
Suppose there is some threshold after which a person is considered old, and before which they are young.

Let Olaf be an old person, and let Yuan be a young person.

1. Suppose I am older than Olaf. Am I old?
2. Suppose I am younger than Olaf. Am I old?
3. Suppose I am older than Yuan. Am I young?
4. Suppose I am younger than Yuan. Am I young?

? 4

Below are graphs of two sequences with positive terms. Assume the sequences continue as shown. Fill in the table with conclusions that can be made from the direct comparison test, if any.



	if $\sum a_n$ converges	if $\sum a_n$ diverges
and if $\{a_n\}$ is the red series	then $\sum b_n$ _____	then $\sum b_n$ _____
and if $\{a_n\}$ is the blue series	then $\sum b_n$ _____	then $\sum b_n$ _____

? 5

For each pair of series below, decide whether the second series is a valid comparison series to determine the convergence of the first series, using the *direct* comparison test and/or the *limit* comparison test.

1. $\sum_{n=10}^{\infty} \frac{1}{n-1}$, compared to the divergent series $\sum_{n=10}^{\infty} \frac{1}{n}$.
2. $\sum_{n=1}^{\infty} \frac{\sin n}{n^2+1}$, compared to the convergent series $\sum_{n=1}^{\infty} \frac{1}{n^2}$.
3. $\sum_{n=5}^{\infty} \frac{n^3+5n+1}{n^6-2}$, compared to the convergent series $\sum_{n=5}^{\infty} \frac{1}{n^3}$.
4. $\sum_{n=5}^{\infty} \frac{1}{\sqrt{n}}$, compared to the divergent series $\sum_{n=5}^{\infty} \frac{1}{\sqrt[4]{n}}$.

? 6

Suppose a_n is a sequence with $\lim_{n \rightarrow \infty} a_n = \frac{1}{2}$. Does $\sum_{n=7}^{\infty} a_n$ converge or diverge, or is it not possible to determine this from the information given? Why?

? 7

What flaw renders the following reasoning invalid?

Q: Determine whether $\sum_{n=1}^{\infty} \frac{\sin n}{n}$ converges or diverges.

A: First, we will evaluate $\lim_{n \rightarrow \infty} \frac{\sin n}{n}$.

- Note $\frac{-1}{n} \leq \frac{\sin n}{n} \leq \frac{1}{n}$ for $n \geq 1$.
- Note also that $\lim_{n \rightarrow \infty} \frac{-1}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} = 0$.
- Therefore, by the Squeeze Theorem, $\lim_{n \rightarrow \infty} \frac{\sin n}{n} = 0$ as well.

So, by the divergence test, $\sum_{n=1}^{\infty} \frac{\sin n}{n}$ converges.

? 8

What flaw renders the following reasoning invalid?

Q: Determine whether $\sum_{n=1}^{\infty} (\sin(\pi n) + 2)$ converges or diverges.

A: We use the integral test. Let $f(x) = \sin(\pi x) + 2$. Note $f(x)$ is always positive, since $\sin(x) + 2 \geq -1 + 2 = 1$. Also, $f(x)$ is continuous.

$$\begin{aligned} \int_1^{\infty} [\sin(\pi x) + 2] dx &= \lim_{b \rightarrow \infty} \int_1^b [\sin(\pi x) + 2] dx \\ &= \lim_{b \rightarrow \infty} \left[-\frac{1}{\pi} \cos(\pi x) + 2x \right]_1^b \\ &= \lim_{b \rightarrow \infty} \left[-\frac{1}{\pi} \cos(\pi b) + 2b + \frac{1}{\pi}(-1) - 2 \right] \\ &= \infty \end{aligned}$$

By the integral test, since the integral diverges, also $\sum_{n=1}^{\infty} (\sin(\pi n) + 2)$ diverges.

? 9

What flaw renders the following reasoning invalid?

Q: Determine whether the series $\sum_{n=1}^{\infty} \frac{2^{n+1}n^2}{e^n + 2n}$ converges or diverges.

A: We want to compare this series to the series $\sum_{n=1}^{\infty} \frac{2^{n+1}}{e^n}$. Note both this series and the series in the question have positive terms.

First, we find that $\frac{2^{n+1}n^2}{e^n + 2n} > \frac{2^{n+1}}{e^n}$ when n is sufficiently large. The justification for this claim is as follows:

- We note that $e^n(n^2 - 1) > n^2 - 1 > 2n$ for n sufficiently large.
- Therefore, $e^n \cdot n^2 > e^n + 2n$
- Therefore, $2^{n+1} \cdot e^n \cdot n^2 > 2^{n+1}(e^n + 2n)$

- Since $e^n + 2n$ and e^n are both expressions that work out to be positive for the values of n under consideration, we can divide both sides of the inequality by these terms without having to flip the inequality. So, $\frac{2^{n+1}n^2}{e^n + 2n} > \frac{2^{n+1}}{e^n}$.

Now, we claim $\sum_{n=1}^{\infty} \frac{2^{n+1}}{e^n}$ converges.

Note $\sum_{n=1}^{\infty} \frac{2^{n+1}}{e^n} = 2 \sum_{n=1}^{\infty} \frac{2^n}{e^n} = 2 \sum_{n=1}^{\infty} \left(\frac{2}{e}\right)^n$. This is a geometric series with $r = \frac{2}{e}$.

Since $2/e < 1$, the series converges.

Now, by the Direct Comparison Test, we conclude that $\sum_{n=1}^{\infty} \frac{2^{n+1}n^2}{e^n + 2n}$ converges.

? 10

Which of the series below are alternating?

(A) $\sum_{n=1}^{\infty} \sin n$

(B) $\sum_{n=1}^{\infty} \frac{\cos(\pi n)}{n^3}$

(C) $\sum_{n=1}^{\infty} \frac{7}{(-n)^{2n}}$

(D) $\sum_{n=1}^{\infty} \frac{(-2)^n}{3^{n+1}}$

? 11

Give an example of a convergent series for which the ratio test is inconclusive.

? 12

Imagine you're taking an exam, and you momentarily forget exactly how the inequality in the ratio test works. You remember there's a ratio, but you don't remember which term goes on top; you remember there's something about the limit being greater than or less than one, but you don't remember which way implies convergence.

Explain why

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| > 1$$

or, equivalently,

$$\lim_{n \rightarrow \infty} \left| \frac{a_n}{a_{n+1}} \right| < 1$$

should mean that the sum $\sum_{n=1}^{\infty} a_n$ diverges (rather than converging).

? 13

Give an example of a series $\sum_{n=a}^{\infty} a_n$, with a function $f(x)$ such that $f(n) = a_n$ for all whole numbers n , such that:

- $\int_a^{\infty} f(x) dx$ diverges, while
- $\sum_{n=a}^{\infty} a_n$ converges.

? 14 (*)

Suppose that you want to use the Limit Comparison Test on the series $\sum_{n=0}^{\infty} a_n$ where $a_n = \frac{2^n + n}{3^n + 1}$.

Write down a sequence $\{b_n\}$ such that $\lim_{n \rightarrow \infty} \frac{a_n}{b_n}$ exists and is nonzero. (You don't have to carry out the Limit Comparison Test)

? 15 (*)

Decide whether each of the following statements is true or false. If false, provide a counterexample. If true provide a brief justification.

1. If $\lim_{n \rightarrow \infty} a_n = 0$, then $\sum_{n=1}^{\infty} a_n$ converges.
2. If $\lim_{n \rightarrow \infty} a_n = 0$, then $\sum_{n=1}^{\infty} (-1)^n a_n$ converges.
3. If $0 \leq a_n \leq b_n$ and $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.

Stage 2

? 16 (*)

Does the series $\sum_{n=2}^{\infty} \frac{n^2}{3n^2 + \sqrt{n}}$ converge?

? 17 (*)

Determine, with explanation, whether the series $\sum_{n=1}^{\infty} \frac{5^k}{4^k + 3^k}$ converges or diverges.

? 18 (*)

Determine whether the series $\sum_{n=0}^{\infty} \frac{1}{n + \frac{1}{2}}$ is convergent or divergent. If it is convergent, find its value.

? 19 (*)

Does the following series converge or diverge? $\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}\sqrt{k+1}}$

? 20

Evaluate the following series, or show that it diverges: $\sum_{k=30}^{\infty} 3(1.001)^k$.

? 21

Evaluate the following series, or show that it diverges: $\sum_{n=3}^{\infty} \left(\frac{-1}{5}\right)^n$.

? 22

Does the following series converge or diverge? $\sum_{n=7}^{\infty} \sin(\pi n)$

? 23

Does the following series converge or diverge? $\sum_{n=7}^{\infty} \cos(\pi n)$

? 24

Does the following series converge or diverge? $\sum_{k=1}^{\infty} \frac{e^k}{k!}$.

? 25

Evaluate the following series, or show that it diverges: $\sum_{k=0}^{\infty} \frac{2^k}{3^{k+2}}$.

? 26

Does the following series converge or diverge? $\sum_{n=1}^{\infty} \frac{n!n!}{(2n)!}$.

? 27

Does the following series converge or diverge? $\sum_{n=1}^{\infty} \frac{n^2 + 1}{2n^4 + n}$.

? 28 (*)

Show that the series $\sum_{n=3}^{\infty} \frac{5}{n(\log n)^{3/2}}$ converges.

? 29 (*)

Find the values of p for which the series $\sum_{n=2}^{\infty} \frac{1}{n(\log n)^p}$ converges.

? 30 (*)

Does $\sum_{n=1}^{\infty} \frac{e^{-\sqrt{n}}}{\sqrt{n}}$ converge or diverge?

? 31 (*)

Use the comparison test (not the limit comparison test) to show whether the series

$$\sum_{n=2}^{\infty} \frac{\sqrt{3n^2 - 7}}{n^3}$$

converges or diverges.

? 32 (*)

Determine whether the series $\sum_{k=1}^{\infty} \frac{\sqrt[3]{k^4 + 1}}{\sqrt{k^5 + 9}}$ converges.

? 33 (*)

Does $\sum_{n=1}^{\infty} \frac{n^4 2^{n/3}}{(2n+7)^4}$ converge or diverge?

? 34 (*)

Determine, with explanation, whether each of the following series converge or diverge.

1. $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n^2 + 1}}$
2. $\sum_{n=1}^{\infty} \frac{n \cos(n\pi)}{2^n}$

? 35 (*)

Determine whether the series

$$\sum_{k=1}^{\infty} \frac{k^4 - 2k^3 + 2}{k^5 + k^2 + k}$$

converges or diverges.

? 36 (*)

Determine whether each of the following series converge or diverge.

1. $\sum_{n=2}^{\infty} \frac{n^2 + n + 1}{n^5 - n}$
2. $\sum_{m=1}^{\infty} \frac{3m + \sin \sqrt{m}}{m^2}$

? 37

Evaluate the following series, or show that it diverges: $\sum_{n=5}^{\infty} \frac{1}{e^n}$.

? 38 (*)

Determine whether the series $\sum_{n=2}^{\infty} \frac{6}{7^n}$ is convergent or divergent. If it is convergent, find its value.

? 39 (*)

Determine, with explanation, whether each of the following series converge or diverge.

1. $1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{9} + \dots$
2. $\sum_{n=1}^{\infty} \frac{2n+1}{2^{2n+1}}$

? 40 (*)

Determine, with explanation, whether each of the following series converges or diverges.

1. $\sum_{k=2}^{\infty} \frac{\sqrt[3]{k}}{k^2 - k}$
2. $\sum_{k=1}^{\infty} \frac{k^{10} 10^k (k!)^2}{(2k)!}$
3. $\sum_{k=3}^{\infty} \frac{1}{k(\log k)(\log \log k)}$

? 41 (*)

Determine whether the series $\sum_{n=1}^{\infty} \frac{n^3 - 4}{2n^5 - 6n}$ is convergent or divergent.

? 42 (*)

What is the smallest value of N such that the partial sum $\sum_{n=1}^N \frac{(-1)^n}{n \cdot 10^n}$ approximates $\sum_{n=1}^{\infty} \frac{(-1)^n}{n \cdot 10^n}$ within an accuracy of 10^{-6} ?

? 43 (*)

It is known that $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^2} = \frac{\pi^2}{12}$ (you don't have to show this). Find N so that S_N , the N^{th} partial sum of the series, satisfies $|\frac{\pi^2}{12} - S_N| \leq 10^{-6}$. Be sure to say why your method can be applied to this particular series.

? 44 (*)

The series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(2n+1)^2}$ converges to some number S (you don't have to prove this). According to the Alternating Series Estimation Theorem, what is the smallest value of N for which the N^{th} partial sum of the series is at most $\frac{1}{100}$ away from S ? For this value of N , write out the N^{th} partial sum of the series.

Stage 3

A number of phenomena roughly follow a distribution called Zipf's law. We discuss some of these in Questions 52 and 53.

? 45 (*)

Determine, with explanation, whether the following series converge or diverge.

1. $\sum_{n=1}^{\infty} \frac{n^n}{9^n n!}$
2. $\sum_{n=1}^{\infty} \frac{1}{n^{\log n}}$

? 46 (*)

(a) Prove that $\int_2^{\infty} \frac{x + \sin x}{1 + x^2} dx$ diverges.

(b) Explain why you *cannot* conclude that $\sum_{n=1}^{\infty} \frac{n + \sin n}{1 + n^2}$ diverges from part (a) and the Integral Test.

(c) Determine, with explanation, whether $\sum_{n=1}^{\infty} \frac{n + \sin n}{1 + n^2}$ converges or diverges.

? 47 (*)

Show that $\sum_{n=1}^{\infty} \frac{e^{-\sqrt{n}}}{\sqrt{n}}$ converges and find an interval of length 0.05 or less that contains its exact value.

? 48 (*)

Suppose that the series $\sum_{n=1}^{\infty} a_n$ converges and that $1 > a_n \geq 0$ for all n . Prove that the series $\sum_{n=1}^{\infty} \frac{a_n}{1 - a_n}$ also converges.

? 49 (*)

Suppose that the series $\sum_{n=0}^{\infty} (1 - a_n)$ converges, where $a_n > 0$ for $n = 0, 1, 2, 3, \dots$. Determine whether the series $\sum_{n=0}^{\infty} 2^n a_n$ converges or diverges.

? 50 (*)

Assume that the series $\sum_{n=1}^{\infty} \frac{na_n - 2n + 1}{n + 1}$ converges, where $a_n > 0$ for $n = 1, 2, \dots$. Is the following series

$$-\log a_1 + \sum_{n=1}^{\infty} \log \left(\frac{a_n}{a_{n+1}} \right)$$

convergent? If your answer is NO, justify your answer. If your answer is YES, evaluate the sum of the series $-\log a_1 + \sum_{n=1}^{\infty} \log \left(\frac{a_n}{a_{n+1}} \right)$.

? 51 (*)

Prove that if $a_n \geq 0$ for all n and if the series $\sum_{n=1}^{\infty} a_n$ converges, then the series $\sum_{n=1}^{\infty} a_n^2$ also converges.

? 52

Suppose the frequency of word use in a language has the following pattern:

The n -th most frequently used word accounts for $\frac{\alpha}{n}$ percent of the total words used.

So, in a text of 100 words, we expect the most frequently used word to appear α times, while the second-most-frequently used word should appear about $\frac{\alpha}{2}$ times, and so on.

If books written in this language use 20,000 distinct words, then the most commonly used word accounts for roughly what percentage of total words used?

? 53

Suppose the sizes of cities in a country adhere to the following pattern: if the largest city has population α , then the n -th largest city has population $\frac{\alpha}{n}$.

If the largest city in this country has 2 million people and the smallest city has 1 person, then the population of the entire country is $\sum_{n=1}^{2 \times 10^6} \frac{2 \times 10^6}{n}$. (For many n 's in this sum $\frac{2 \times 10^6}{n}$ is not an integer. Ignore that.) Evaluate this sum approximately, with an error of no more than 1 million people.

- The authors should be a little more careful making such a blanket statement. While it is true that it is not wise to approximate a divergent series by taking N terms with N large, there are cases when one can get a very good approximation by taking N terms with N small! For example, the Taylor remainder theorem shows us that when the n^{th} derivative of a function $f(x)$ grows very quickly with n , Taylor polynomials of degree N , with N large, can give bad approximations of $f(x)$, while the Taylor polynomials of degree one or two can still provide very good approximations of $f(x)$ when x is very small. As an example of this, one of the triumphs of quantum electrodynamics, namely the computation of the anomalous magnetic moment of the electron, depends on precisely this. A number of important quantities were predicted using the first few terms of divergent power series. When those quantities were measured experimentally, the predictions turned out to be incredibly accurate.
- The field of asymptotic analysis often makes use of the first few terms of divergent series to generate approximate solutions to problems; this, along with numerical computations, is one of the most important techniques in applied mathematics. Indeed, there is a whole wonderful book (which, unfortunately, is too advanced for most Calculus 2 students) devoted to playing with divergent series called, unsurprisingly, “Divergent Series” by G.H. Hardy. This is not to be confused with the “Divergent” series by V. Roth set in a post-apocalyptic dystopian Chicago. That latter series diverges quite dramatically from mathematical topics, while the former does not have a film adaptation (yet).
- We have discussed the contrapositive a few times in the CLP notes, but it doesn't hurt to discuss it again here (or for the reader to quickly look up the relevant footnote in Section 1.3 of the CLP-1 text). At any rate, given a statement of the form “If A is true, then B is true” the contrapositive is “If B is not true, then A is not true”. The two statements in quotation marks are logically equivalent — if one is true, then so is the other. In the present context we have “If $(\sum a_n \text{ converges})$ then $(a_n \text{ converges to } 0)$.” The contrapositive of this statement is then “If $(a_n \text{ does not converge to } 0)$ then $(\sum a_n \text{ does not converge})$.”
- This series converges to Apéry's constant $1.2020569031 \dots$. The constant is named for Roger Apéry (1916–1994) who proved that this number must be irrational. This number appears in many contexts including the following cute fact — the reciprocal of Apéry's constant gives the probability that three positive integers, chosen at random, do not share a common prime factor.
- Latin for “Once the necessary changes are made”. This phrase still gets used a little, but these days mathematicians tend to write something equivalent in English. Indeed, English is pretty much the *lingua franca* for mathematical publishing. *Quidquid erit*.
- This series, viewed as a function of p , is called the Riemann zeta function, $\zeta(p)$, or the Euler-Riemann zeta function. It is extremely important because of its connections to prime numbers (among many other things). Indeed Euler proved that $\zeta(p) = \sum_{n=1}^{\infty} \frac{1}{n^p} = \prod_{p \text{ prime}} (1 - p^{-p})^{-1} \dots$ Riemann showed the connections between the zeros of this function (over

complex numbers p) and the distribution of prime numbers. Arguably the most famous unsolved problem in mathematics, the Riemann hypothesis, concerns the locations of zeros of this function.

7. We could go even further and see what happens if we include powers of $\log(\log(n))$ and other more exotic slow growing functions.
8. Go back and quickly scan Theorem 3.3.5; to apply it we need to show that $\frac{1}{n^2+2n+3}$ is positive and decreasing (it is), and then we need to integrate $\int \frac{1}{x^2+2x+3} dx$. To do that we reread the notes on partial fractions, then rewrite $x^2 + 2x + 3 = (x + 1)^2 + 2$ and so $\int_1^\infty \frac{1}{x^2+2x+3} dx = \int_1^\infty \frac{1}{(x+1)^2+2} dx \cdots$ and then arctangent appears, etc etc. Urgh. Okay — let's go back to the text now and see how to avoid this.
9. To understand this consider any series $\sum_{n=1}^\infty a_n$. We can always cut such a series into two parts — pick some huge number like 10^6 . Then $\sum_{n=1}^\infty a_n = \sum_{n=1}^{10^6} a_n + \sum_{n=10^6+1}^\infty a_n$. The first sum, though it could be humongous, is finite. So the left hand side, $\sum_{n=1}^\infty a_n$, is a well-defined finite number if and only if $\sum_{n=10^6+1}^\infty a_n$, is a well-defined finite number. The convergence or divergence of the series is determined by the second sum, which only contains a_n for “large” n .
10. The symbol “ \gg ” means “much larger than”. Similarly, the symbol “ \ll ” means “much less than”. Good shorthand symbols can be quite expressive.
11. This is very similar to how we computed limits at infinity way way back near the beginning of CLP-1.
12. Really, really tedious. And you thought some of those partial fractions computations were bad ...
13. The interested reader may wish to check out “Stirling's approximation”, which says that $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$.
14. We shall see later, in Theorem 3.5.13, that the function $\sum_{n=0}^\infty a_n x^{n-1}$ is indeed the derivative of the function $\sum_{n=0}^\infty a_n x^n$. Of course, such a statement only makes sense where these series converge — how can you differentiate a divergent series? (This is not an allusion to a popular series of dystopian novels.) Actually, there is quite a bit of interesting and useful mathematics involving divergent series, but it is well beyond the scope of this course.
15. We shall also see later, in Theorem 3.5.13, that the function $\sum_{n=0}^\infty \frac{a}{n+1} x^{n+1}$ is indeed an antiderivative of the function $\sum_{n=0}^\infty a x^n$.
16. It might be a good idea to review the beginning of §2.3 at this point.
17. At least if our table is strong enough.
18. The interested reader should use their favourite search engine to read more on the link between this series and musical harmonics. You can also find interesting links between the Harmonic series and the so-called “jeep problem” and also the problem of stacking a tower of dominoes to create an overhang that does not topple over.
19. The grouping argument can be generalised further and the interested reader should look up Cauchy's condensation test.
20. Knowing that the left and right hand sides of (\star) are zero for the same values of x is far from the end of the story. Two functions $f(x)$ and $g(x)$ having the same zeros, need not be equal. It is certainly possible that $f(x) = g(x) * A(x)$ where $A(x)$ is a function that is nowhere zero. The interested reader should look up the Weierstrass factorisation theorem.
21. Search-engine your way to “Riemann zeta function”.
22. It is one way to state a property of the real number system called “completeness”. The interested reader should use their favourite search engine to look up “completeness of the real numbers”.

This page titled [3.3: Convergence Tests](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

3.4: Absolute and Conditional Convergence

We have now seen examples of series that converge and of series that diverge. But we haven't really discussed how robust the convergence of series is — that is, can we tweak the coefficients in some way while leaving the convergence unchanged. A good example of this is the series

$$\sum_{n=1}^{\infty} \left(\frac{1}{3}\right)^n$$

This is a simple geometric series and we know it converges. We have also seen, as examples 3.3.20 and 3.3.21 showed us, that we can multiply or divide the n^{th} term by n and it will still converge. We can even multiply the n^{th} term by $(-1)^n$ (making it an alternating series), and it will still converge. Pretty robust.

On the other hand, we have explored the Harmonic series and its relatives quite a lot and we know it is much more delicate. While

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverges, we also know the following two series converge:

$$\sum_{n=1}^{\infty} \frac{1}{n^{1.00000001}} \quad \sum_{n=1}^{\infty} (-1)^n \frac{1}{n}.$$

This suggests that the divergence of the Harmonic series is much more delicate. In this section, we discuss one way to characterise this sort of delicate convergence — especially in the presence of changes of sign.

Definitions

Definition 3.4.1 Absolute and conditional convergence

1. A series $\sum_{n=1}^{\infty} a_n$ is said to converge absolutely if the series $\sum_{n=1}^{\infty} |a_n|$ converges.
2. If $\sum_{n=1}^{\infty} a_n$ converges but $\sum_{n=1}^{\infty} |a_n|$ diverges we say that $\sum_{n=1}^{\infty} a_n$ is conditionally convergent.

If you consider these definitions for a moment, it should be clear that absolute convergence is a stronger condition than just simple convergence. All the terms in $\sum_n |a_n|$ are forced to be positive (by the absolute value signs), so that $\sum_n |a_n|$ must be bigger than $\sum_n a_n$ — making it easier for $\sum_n |a_n|$ to diverge. This is formalised by the following theorem, which is an immediate consequence of the comparison test, Theorem 3.3.8.a, with $c_n = |a_n|$.

Theorem 3.4.2 Absolute convergence implies convergence

If the series $\sum_{n=1}^{\infty} |a_n|$ converges then the series $\sum_{n=1}^{\infty} a_n$ also converges. That is, absolute convergence implies convergence.

Recall that some of our convergence tests (for example, the integral test) may only be applied to series with positive terms. Theorem 3.4.2 opens up the possibility of applying “positive only” convergence tests to series whose terms are not all positive, by checking for “absolute convergence” rather than for plain “convergence”.

Example 3.4.3 $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$

The alternating harmonic series $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$ of Example 3.3.15 converges (by the alternating series test). But the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ of Example 3.3.6 diverges (by the integral test). So the alternating harmonic series $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$ converges conditionally.

✓ Example 3.4.4 $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n^2}$

Because the series $\sum_{n=1}^{\infty} |(-1)^{n-1} \frac{1}{n^2}| = \sum_{n=1}^{\infty} \frac{1}{n^2}$ of Example 3.3.6 converges (by the integral test), the series $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n^2}$ converges absolutely, and hence converges.

✓ Example 3.4.5 Random signs

Imagine flipping a coin infinitely many times. Set $\sigma_n = +1$ if the n^{th} flip comes up heads and $\sigma_n = -1$ if the n^{th} flip comes up tails. The series $\sum_{n=1}^{\infty} (-1)^{\sigma_n} \frac{1}{n^2}$ is not in general an alternating series. But we know that the series $\sum_{n=1}^{\infty} |(-1)^{\sigma_n} \frac{1}{n^2}| = \sum_{n=1}^{\infty} \frac{1}{n^2}$ converges. So $\sum_{n=1}^{\infty} (-1)^{\sigma_n} \frac{1}{n^2}$ converges absolutely, and hence converges.

Optional — The delicacy of conditionally convergent series

Conditionally convergent series have to be treated with great care. For example, switching the order of the terms in a finite sum does not change its value.

$$1 + 2 + 3 + 4 + 5 + 6 = 6 + 3 + 5 + 2 + 4 + 1$$

The same is true for absolutely convergent series. But it is *not true* for conditionally convergent series. In fact by reordering *any* conditionally convergent series, you can make it add up to *any* number you like, including $+\infty$ and $-\infty$. This very strange result is known as Riemann's rearrangement theorem, named after Bernhard Riemann (1826–1866). The following example illustrates the phenomenon.

✓ Example 3.4.6 The alternating Harmonic series

The alternating Harmonic series

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$$

is a very good example of conditional convergence. We can show, quite explicitly, how we can rearrange the terms to make it add up to two different numbers. Later, in Example 3.5.20, we'll show that this series is equal to $\log 2$. However, by rearranging the terms we can make it sum to $\frac{1}{2} \log 2$. The usual order is

$$\frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \dots$$

For the moment think of the terms being paired as follows:

$$\left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{5} - \frac{1}{6}\right) + \dots$$

so the denominators go odd-even odd-even. Now rearrange the terms so the denominators are odd-even-even odd-even-even:

$$\left(1 - \frac{1}{2} - \frac{1}{4}\right) + \left(\frac{1}{3} - \frac{1}{6} - \frac{1}{8}\right) + \left(\frac{1}{5} - \frac{1}{10} - \frac{1}{12}\right) + \dots$$

Now notice that the first term in each triple is exactly twice the second term. If we now combine those terms we get

$$\begin{aligned} & \left(\underbrace{1 - \frac{1}{2} - \frac{1}{4}}_{=1/2}\right) + \left(\underbrace{\frac{1}{3} - \frac{1}{6} - \frac{1}{8}}_{=1/6}\right) + \left(\underbrace{\frac{1}{5} - \frac{1}{10} - \frac{1}{12}}_{=1/10}\right) + \dots \\ &= \left(\frac{1}{2} - \frac{1}{4}\right) + \left(\frac{1}{6} - \frac{1}{8}\right) + \left(\frac{1}{10} - \frac{1}{12}\right) + \dots \end{aligned}$$

We can now extract a factor of $\frac{1}{2}$ from each term, so

$$\begin{aligned} &= \frac{1}{2} \left(\frac{1}{1} - \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{3} - \frac{1}{4} \right) + \frac{1}{2} \left(\frac{1}{5} - \frac{1}{6} \right) + \dots \\ &= \frac{1}{2} \left[\left(\frac{1}{1} - \frac{1}{2} \right) + \left(\frac{1}{3} - \frac{1}{4} \right) + \left(\frac{1}{5} - \frac{1}{6} \right) + \dots \right] \end{aligned}$$

So by rearranging the terms, the sum of the series is now exactly half the original sum!

In fact, we can go even further, and show how we can rearrange the terms of the alternating harmonic series to add up to any given number¹. For the purposes of the example we have chosen 1.234, but it could really be any number. The example below can actually be formalised to give a proof of the rearrangement theorem.

✓ Example 3.4.7 Reorder summands to get 1.234

We'll show how to reorder the conditionally convergent series $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$ so that it adds up to exactly 1.234 (but the reader should keep in mind that any fixed number will work).

- First create two lists of numbers — the first list consisting of the positive terms of the series, in order, and the second consisting of the negative numbers of the series, in order.

$$1, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \dots \quad \text{and} \quad -\frac{1}{2}, -\frac{1}{4}, -\frac{1}{6}, \dots$$

- Notice that that if we add together the numbers in the second list, we get

$$-\frac{1}{2} \left[1 + \frac{1}{2} + \frac{1}{3} + \dots \right]$$

which is just $-\frac{1}{2}$ times the harmonic series. So the numbers in the second list add up to $-\infty$.

Also, if we add together the numbers in the first list, we get

$$1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} \dots \quad \text{which is greater than} \quad \frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \dots$$

That is, the sum of the first set of numbers must be bigger than the sum of the second set of numbers (which is just -1 times the second list). So the numbers in the first list add up to $+\infty$.

- Now we build up our reordered series. Start by moving just enough numbers from the beginning of the first list into the reordered series to get a sum bigger than 1.234.

$$1 + \frac{1}{3} = 1.3333$$

We know that we can do this, because the sum of the terms in the first list diverges to $+\infty$.

- Next move just enough numbers from the beginning of the second list into the reordered series to get a number less than 1.234.

$$1 + \frac{1}{3} - \frac{1}{2} = 0.8333$$

Again, we know that we can do this because the sum of the numbers in the second list diverges to $-\infty$.

- Next move just enough numbers from the beginning of the remaining part of the first list into the reordered series to get a number bigger than 1.234.

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} + \frac{1}{9} = 1.2873$$

Again, this is possible because the sum of the numbers in the first list diverges. Even though we have already used the first few numbers, the sum of the rest of the list will still diverge.

- Next move just enough numbers from the beginning of the remaining part of the second list into the reordered series to get a number less than 1.234.

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} + \frac{1}{9} - \frac{1}{4} = 1.0373$$

- At this point the idea is clear, just keep going like this. At the end of each step, the difference between the sum and 1.234 is smaller than the magnitude of the first unused number in the lists. Since the numbers in both lists tend to zero as you go farther and farther up the list, this procedure will generate a series whose sum is exactly 1.234. Since in each step we remove at least one number from a list and we alternate between the two lists, the reordered series will contain all of the terms from $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$, with each term appearing exactly once.

Exercises

Stage 1

? 1 (*)

Decide whether the following statement is true or false. If false, provide a counterexample. If true provide a brief justification.

- If $\sum_{n=1}^{\infty} (-1)^{n+1} b_n$ converges, then $\sum_{n=1}^{\infty} b_n$ also converges.

? 2

Describe the series $\sum_{n=1}^{\infty} a_n$ based on whether $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} |a_n|$ converge or diverge, using vocabulary from this section where possible.

	$\sum a_n$ converges	$\sum a_n$ diverges
$\sum a_n $ converges		
$\sum a_n $ diverges		

Stage 2

? 3 (*)

Determine whether the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{9n+5}$ is absolutely convergent, conditionally convergent, or divergent; justify your answer.

? 4 (*)

Determine whether the series $\sum_{n=1}^{\infty} \frac{(-1)^{2n+1}}{1+n}$ is absolutely convergent, conditionally convergent, or divergent.

? 5 (*)

The series $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1+4^n}{3+2^{2n}}$ either:

- converges absolutely;
- converges conditionally;

- diverges;
- or none of the above.

Determine which is correct.

? 6 (*)

Does the series $\sum_{n=5}^{\infty} \frac{\sqrt{n} \cos n}{n^2 - 1}$ converge conditionally, converge absolutely, or diverge?

? 7 (*)

Determine (with justification!) whether the series $\sum_{n=1}^{\infty} \frac{n^2 - \sin n}{n^6 + n^2}$ converges absolutely, converges but not absolutely, or diverges.

? 8 (*)

Determine (with justification!) whether the series $\sum_{n=0}^{\infty} \frac{(-1)^n (2n)!}{(n^2 + 1)(n!)^2}$ converges absolutely, converges but not absolutely, or diverges.

? 9 (*)

Determine (with justification!) whether the series $\sum_{n=2}^{\infty} \frac{(-1)^n}{n(\log n)^{101}}$ converges absolutely, converges but not absolutely, or diverges.

? 10

Show that the series $\sum_{n=1}^{\infty} \frac{\sin n}{n^2}$ converges.

? 11

Show that the series $\sum_{n=1}^{\infty} \left(\frac{\sin n}{4} - \frac{1}{8} \right)^n$ converges.

? 12

Show that the series $\sum_{n=1}^{\infty} \frac{\sin^2 n - \cos^2 n + \frac{1}{2}}{2^n}$ converges.

Stage 3

? 13 (*)

Both parts of this question concern the series $S = \sum_{n=1}^{\infty} (-1)^{n-1} 24n^2 e^{-n^3}$.

1. Show that the series S converges absolutely.
2. Suppose that you approximate the series S by its fifth partial sum S_5 . Give an upper bound for the error resulting from this approximation.

? 14

You may assume without proof the following:

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} = \cos(1)$$

Using this fact, approximate $\cos 1$ as a rational number, accurate to within $\frac{1}{1000}$.

Check your answer against a calculator's approximation of $\cos(1)$: what was your actual error?

? 15

Let a_n be defined as

$$a_n = \begin{cases} -e^{n/2} & \text{if } n \text{ is prime} \\ n^2 & \text{if } n \text{ is not prime} \end{cases}$$

Show that the series $\sum_{n=1}^{\infty} \frac{a_n}{e^n}$ converges.

1. This is reminiscent of the accounting trick of pushing all the company's debts off to next year so that this year's accounts look really good and you can collect your bonus.

This page titled [3.4: Absolute and Conditional Convergence](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

3.5: Power Series

Let's return to the simple geometric series

$$\sum_{n=0}^{\infty} x^n$$

where x is some real number. As we have seen (back in Example 3.2.4 and Lemma 3.2.5), for $|x| < 1$ this series converges to a limit, that varies with x , while for $|x| \geq 1$ the series diverges. Consequently we can consider this series to be a function of x

$$f(x) = \sum_{n=0}^{\infty} x^n \quad \text{on the domain } |x| < 1.$$

Furthermore (also from Example 3.2.4 and Lemma 3.2.5) we know what the function is.

$$f(x) = \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}.$$

Hence we can consider the series $\sum_{n=0}^{\infty} x^n$ as a new way of representing the function $\frac{1}{1-x}$ when $|x| < 1$. This series is an example of a power series.

Of course, representing a function as simple as $\frac{1}{1-x}$ by a series doesn't seem like it is going to make life easier. However the idea of representing a function by a series turns out to be extremely helpful. Power series turn out to be very robust mathematical objects and interact very nicely with not only standard arithmetic operations, but also with differentiation and integration (see Theorem 3.5.13). This means, for example, that

$$\begin{aligned} \frac{d}{dx} \left\{ \frac{1}{1-x} \right\} &= \frac{d}{dx} \sum_{n=0}^{\infty} x^n && \text{provided } |x| < 1 \\ &= \sum_{n=0}^{\infty} \frac{d}{dx} x^n && \text{just differentiate term by term} \\ &= \sum_{n=0}^{\infty} n x^{n-1} \end{aligned}$$

and in a very similar way

$$\begin{aligned} \int \frac{1}{1-x} dx &= \int \sum_{n=0}^{\infty} x^n dx && \text{provided } |x| < 1 \\ &= \sum_{n=0}^{\infty} \int x^n dx && \text{just integrate term by term} \\ &= C + \sum_{n=0}^{\infty} \frac{1}{n+1} x^{n+1} \end{aligned}$$

We are hiding some mathematics under the word “just” in the above, but you can see that once we have a power series representation of a function, differentiation and integration become very straightforward.

So we should set as our goal for this section, the development of machinery to define and understand power series. This will allow us to answer questions¹ like

$$\text{Is } e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} ?$$

Our starting point (now that we have equipped ourselves with basic ideas about series), is the definition of power series.

Definitions

 Definition 3.5.1

A series of the form

$$A_0 + A_1(x - c) + A_2(x - c)^2 + A_3(x - c)^3 + \dots = \sum_{n=0}^{\infty} A_n(x - c)^n$$

is called a **power series in $(x - c)$** or a **power series centered on c** . The numbers A_n are called the coefficients of the power series.

One often considers power series centered on $c = 0$ and then the series reduces to

$$A_0 + A_1x + A_2x^2 + A_3x^3 + \dots = \sum_{n=0}^{\infty} A_nx^n$$

For example $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ is the power series with $c = 0$ and $A_n = \frac{1}{n!}$. Typically, as in that case, the coefficients A_n are given fixed numbers, but the “ x ” is to be thought of as a variable. Thus each power series is really a whole family of series — a different series for each value of x .

One possible value of x is $x = c$ and then the series reduces² to

$$\begin{aligned} \sum_{n=0}^{\infty} A_n(x - c)^n \Big|_{x=c} &= \sum_{n=0}^{\infty} A_n(c - c)^n \\ &= \underbrace{A_0}_{n=0} + \underbrace{0}_{n=1} + \underbrace{0}_{n=2} + \underbrace{0}_{n=3} + \dots \end{aligned}$$

and so simply converges to A_0 .

We now know that a power series converges when $x = c$. We can now use our convergence tests to determine for what other values of x the series converges. Perhaps most straightforward is the ratio test. The n^{th} term in the series $\sum_{n=0}^{\infty} A_n(x - c)^n$ is $a_n = A_n(x - c)^n$. To apply the ratio test we need to compute the limit

$$\begin{aligned} \lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| &= \lim_{n \rightarrow \infty} \left| \frac{A_{n+1}(x - c)^{n+1}}{A_n(x - c)^n} \right| \\ &= \lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right| \cdot |x - c| \\ &= |x - c| \cdot \lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right|. \end{aligned}$$

When we do so there are several possible outcomes.

- If the limit of ratios exists and is nonzero

$$\lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right| = A \neq 0,$$

then the ratio test says that the series $\sum_{n=0}^{\infty} A_n(x - c)^n$

- converges when $A \cdot |x - c| < 1$, i.e. when $|x - c| < \frac{1}{A}$, and
- diverges when $A \cdot |x - c| > 1$, i.e. when $|x - c| > \frac{1}{A}$.

Because of this, when the limit exists, the quantity

 Equation 3.5.2 Radius of convergence

$$R = \frac{1}{A} = \left[\lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right| \right]^{-1}$$

is called the *radius of convergence* of the series³.

- If the limit of ratios exists and is zero

$$\lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right| = 0$$

then $\lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right| |x - c| = 0$ for every x and the ratio test tells us that the series $\sum_{n=0}^{\infty} A_n (x - c)^n$ converges for every number x . In this case we say that the series has an infinite radius of convergence.

- If the limit of ratios diverges to $+\infty$

$$\lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right| = +\infty$$

then $\lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right| |x - c| = +\infty$ for every $x \neq c$. The ratio test then tells us that the series $\sum_{n=0}^{\infty} A_n (x - c)^n$ diverges for every number $x \neq c$. As we have seen above, when $x = c$, the series reduces to $A_0 + 0 + 0 + 0 + 0 + \dots$, which of course converges. In this case we say that the series has radius of convergence zero.

- If $\left| \frac{A_{n+1}}{A_n} \right|$ does not approach a limit as $n \rightarrow \infty$, then we learn nothing from the ratio test and we must use other tools to understand the convergence of the series.

All of these possibilities do happen. We give an example of each below. But first, the concept of “radius of convergence” is important enough to warrant a formal definition.

Definition 3.5.3

1. Let $0 < R < \infty$. If $\sum_{n=0}^{\infty} A_n (x - c)^n$ converges for $|x - c| < R$, and diverges for $|x - c| > R$, then we say that the series has radius of convergence R .
2. If $\sum_{n=0}^{\infty} A_n (x - c)^n$ converges for every number x , we say that the series has an infinite radius of convergence.
3. If $\sum_{n=0}^{\infty} A_n (x - c)^n$ diverges for every $x \neq c$, we say that the series has radius of convergence zero.

Example 3.5.4 Finite nonzero radius of convergence

We already know that, if $a \neq 0$, the geometric series $\sum_{n=0}^{\infty} ax^n$ converges when $|x| < 1$ and diverges when $|x| \geq 1$. So, in the terminology of Definition 3.5.3, the geometric series has radius of convergence $R = 1$. As a consistency check, we can also compute R using 3.5.2. The series $\sum_{n=0}^{\infty} ax^n$ has $A_n = a$. So

$$R = \left[\lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right| \right]^{-1} = \left[\lim_{n \rightarrow \infty} 1 \right]^{-1} = 1$$

as expected.

Example 3.5.5 Radius of convergence = $+\infty$

The series $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ has $A_n = \frac{1}{n!}$. So

$$\begin{aligned} \lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right| &= \lim_{n \rightarrow \infty} \frac{\frac{1}{(n+1)!}}{\frac{1}{n!}} \\ &= \lim_{n \rightarrow \infty} \frac{n!}{(n+1)!} \\ &= \lim_{n \rightarrow \infty} \frac{1 \times 2 \times 3 \times \dots \times n}{1 \times 2 \times 3 \times \dots \times n \times (n+1)} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n+1} \\ &= 0 \end{aligned}$$

and $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ has radius of convergence ∞ . It converges for every x .

✓ Example 3.5.6 Radius of convergence = 0

The series $\sum_{n=0}^{\infty} n!x^n$ has $A_n = n!$. So

$$\begin{aligned} \lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right| &= \lim_{n \rightarrow \infty} \frac{(n+1)!}{n!} = \lim_{n \rightarrow \infty} \frac{1 \times 2 \times 3 \times 4 \times \cdots \times n \times (n+1)}{1 \times 2 \times 3 \times 4 \times \cdots \times n} \\ &= \lim_{n \rightarrow \infty} (n+1) \\ &= +\infty \end{aligned}$$

and $\sum_{n=0}^{\infty} n!x^n$ has radius of convergence zero⁴. It converges only for $x = 0$, where it takes the value $0! = 1$.

✓ Example 3.5.7 An awkward series to test

Comparing the series

$$1 + 2x + x^2 + 2x^3 + x^4 + 2x^5 + \cdots$$

to

$$\sum_{n=1}^{\infty} A_n x^n = A_0 + A_1 x + A_2 x^2 + A_3 x^3 + A_4 x^4 + A_5 x^5 + \cdots$$

we see that

$$A_0 = 1 \quad A_1 = 2 \quad A_2 = 1 \quad A_3 = 2 \quad A_4 = 1 \quad A_5 = 2 \quad \cdots$$

so that

$$\frac{A_1}{A_0} = 2 \quad \frac{A_2}{A_1} = \frac{1}{2} \quad \frac{A_3}{A_2} = 2 \quad \frac{A_4}{A_3} = \frac{1}{2} \quad \frac{A_5}{A_4} = 2 \quad \cdots$$

and $\frac{A_{n+1}}{A_n}$ does not converge as $n \rightarrow \infty$. Since the limit of the ratios does not exist, we cannot tell anything from the ratio test. Nonetheless, we can still figure out for which x 's our power series converges.

- Because every coefficient A_n is either 1 or 2, the n^{th} term in our series obeys

$$|A_n x^n| \leq 2|x|^n$$

and so is smaller than the n^{th} term in the geometric series $\sum_{n=0}^{\infty} 2|x|^n$. This geometric series converges if $|x| < 1$. So, by the comparison test, our series converges for $|x| < 1$ too.

- Since every A_n is at least one, the n^{th} term in our series obeys

$$|A_n x^n| \geq |x|^n$$

If $|x| \geq 1$, this $a_n = A_n x^n$ cannot converge to zero as $n \rightarrow \infty$, and our series diverges by the divergence test.

In conclusion, our series converges if and only if $|x| < 1$, and so has radius of convergence 1.

✓ Example 3.5.8 A series from π

Lets construct a series from the digits of π . Now to avoid dividing by zero, let us set

$$A_n = 1 + \text{the } n^{\text{th}} \text{ digit of } \pi$$

Since $\pi = 3.141591\dots$

$$A_0 = 4 \quad A_1 = 2 \quad A_2 = 5 \quad A_3 = 2 \quad A_4 = 6 \quad A_5 = 10 \quad A_6 = 2 \quad \dots$$

Consequently every A_n is an integer between 1 and 10 and gives us the series

$$\sum_{n=0}^{\infty} A_n x^n = 4 + 2x + 5x^2 + 2x^3 + 6x^4 + 10x^5 + \dots$$

The number π is irrational⁵ and consequently the ratio $\frac{A_{n+1}}{A_n}$ cannot have a limit as $n \rightarrow \infty$. If you do not understand why this is the case then don't worry too much about it⁶. As in the last example, the limit of the ratios does not exist and we cannot tell anything from the ratio test. But we can still figure out for which x 's it converges.

- Because every coefficient A_n is no bigger (in magnitude) than 10, the n^{th} term in our series obeys

$$|A_n x^n| \leq 10|x|^n$$

and so is smaller than the n^{th} term in the geometric series $\sum_{n=0}^{\infty} 10|x|^n$. This geometric series converges if $|x| < 1$. So, by the comparison test, our series converges for $|x| < 1$ too.

- Since every A_n is at least one, the n^{th} term in our series obeys

$$|A_n x^n| \geq |x|^n$$

If $|x| \geq 1$, this $a_n = A_n x^n$ cannot converge to zero as $n \rightarrow \infty$, and our series diverges by the divergence test.

In conclusion, our series converges if and only if $|x| < 1$, and so has radius of convergence 1.

Though we won't prove it, it is true that every power series has a radius of convergence, whether or not the limit $\lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right|$ exists.

Theorem 3.5.9

Let $\sum_{n=0}^{\infty} A_n (x - c)^n$ be a power series. Then one of the following alternatives must hold.

1. The power series converges for every number x . In this case we say that the radius of convergence is ∞ .
2. There is a number $0 < R < \infty$ such that the series converges for $|x - c| < R$ and diverges for $|x - c| > R$. Then R is called the radius of convergence.
3. The series converges for $x = c$ and diverges for all $x \neq c$. In this case, we say that the radius of convergence is 0.

Definition 3.5.10

Consider the power series

$$\sum_{n=0}^{\infty} A_n (x - c)^n.$$

The set of real x -values for which it converges is called the interval of convergence of the series.

Suppose that the power series $\sum_{n=0}^{\infty} A_n (x - c)^n$ has radius of convergence R . Then from Theorem 3.5.9, we have that

- if $R = \infty$, then its interval of convergence is $-\infty < x < \infty$, which is also denoted $(-\infty, \infty)$, and
- if $R = 0$, then its interval of convergence is just the point $x = c$, and
- if $0 < R < \infty$, then we know that the series converges for any x which obeys

$$\begin{aligned} |x - c| < R & \quad \text{or equivalently} \quad -R < x - c < R \\ & \quad \text{or equivalently} \quad c - R < x < c + R \end{aligned}$$

But we do not (yet) know whether or not the series converges at the two end points of that interval. We do know, however, that its interval of convergence must be one of

- $c - R < x < c + R$, which is also denoted $(c - R, c + R)$, or

- $c - R \leq x < c + R$, which is also denoted $[c - R, c + R)$, or
- $c - R < x \leq c + R$, which is also denoted $(c - R, c + R]$, or
- $c - R \leq x \leq c + R$, which is also denoted $[c - R, c + R]$.

To reiterate — while the radius convergence, R with $0 < R < \infty$, tells us that the series converges for $|x - c| < R$ and diverges for $|x - c| > R$, it does not (by itself) tell us whether or not the series converges when $|x - c| = R$, i.e. when $x = c \pm R$. The following example shows that all four possibilities can occur.

✓ **Example 3.5.11** The series $\sum_{n=1}^{\infty} \frac{x^n}{n^p}$

Let p be any real number⁷ and consider the series $\sum_{n=1}^{\infty} \frac{x^n}{n^p}$. This series has $A_n = \frac{1}{n^p}$. Since

$$\lim_{n \rightarrow \infty} \left| \frac{A_{n+1}}{A_n} \right| = \lim_{n \rightarrow \infty} \frac{n^p}{(n+1)^p} = \lim_{n \rightarrow \infty} \frac{1}{\left(1 + \frac{1}{n}\right)^p} = 1$$

the series has radius of convergence 1. So it certainly converges for $|x| < 1$ and diverges for $|x| > 1$. That just leaves $x = \pm 1$.

- When $x = 1$, the series reduces to $\sum_{n=1}^{\infty} \frac{1}{n^p}$. We know, from Example 3.3.6, that this series converges if and only if $p > 1$.
- When $x = -1$, the series reduces to $\sum_{n=1}^{\infty} \frac{(-1)^n}{n^p}$. By the alternating series test, Theorem 3.3.14, this series converges whenever $p > 0$ (so that $\frac{1}{n^p}$ tends to zero as n tends to infinity). When $p \leq 0$ (so that $\frac{1}{n^p}$ does not tend to zero as n tends to infinity), it diverges by the divergence test, Theorem 3.3.1.

So

- The power series $\sum_{n=1}^{\infty} x^n$ (i.e. $p = 0$) has interval of convergence $-1 < x < 1$.
- The power series $\sum_{n=1}^{\infty} \frac{x^n}{n}$ (i.e. $p = 1$) has interval of convergence $-1 \leq x < 1$.
- The power series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n} x^n$ (i.e. $p = 1$) has interval of convergence $-1 < x \leq 1$.
- The power series $\sum_{n=1}^{\infty} \frac{x^n}{n^2}$ (i.e. $p = 2$) has interval of convergence $-1 \leq x \leq 1$.

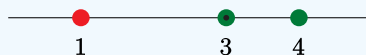
✓ **Example: 3.5.12** Playing with intervals of convergence

We are told that a certain power series with center $c = 3$, converges at $x = 4$ and diverges at $x = 1$. What else can we say about the convergence or divergence of the series for other values of x ?

We are told that the series is centered at 3, so its terms are all powers of $(x - 3)$ and it is of the form

$$\sum_{n \geq 0} A_n (x - 3)^n.$$

A good way to summarise the convergence data we are given is with a figure like the one below. Green dots mark the values of x where the series is known to converge. (Recall that every power series converges at its center.) The red dot marks the value of x where the series is known to diverge. The bull's eye marks the center.



Can we say more about the convergence and/or divergence of the series for other values of x ? Yes!

Let us think about the radius of convergence, R , of the series. We know that it must exist and the information we have been given allows us to bound R . Recall that

- the series converges at x provided that $|x - 3| < R$ and
- the series diverges at x if $|x - 3| > R$.

We have been told that

- the series converges when $x = 4$, which tells us that
 - $x = 4$ cannot obey $|x - 3| > R$ so
 - $x = 4$ must obey $|x - 3| \leq R$, i.e. $|4 - 3| \leq R$, i.e. $R \geq 1$
- the series diverges when $x = 1$ so we also know that

- $x = 1$ cannot obey $|x - 3| < R$ so
- $x = 1$ must obey $|x - 3| \geq R$, i.e. $|1 - 3| \geq R$, i.e. $R \leq 2$

We still don't know R exactly. But we do know that $1 \leq R \leq 2$. Consequently,

- since 1 is the smallest that R could be, the series certainly converges at x if $|x - 3| < 1$, i.e. if $2 < x < 4$ and
- since 2 is the largest that R could be, the series certainly diverges at x if $|x - 3| > 2$, i.e. if $x > 5$ or if $x < 1$.

The following figure provides a resume of all of this convergence data — there is convergence at green x 's and divergence at red x 's.



Notice that from the data given we cannot say anything about the convergence or divergence of the series on the intervals $(1, 2]$ and $(4, 5]$.

One lesson that we can derive from this example is that,

- if a series has center c and converges at a ,
- then it also converges at all points between c and a , as well as at all points of distance strictly less than $|a - c|$ from c on the other side of c from a .

Working With Power Series

Just as we have done previously with limits, differentiation and integration, we can construct power series representations of more complicated functions by using those of simpler functions. Here is a theorem that helps us to do so.

Theorem 3.5.13 Operations on Power Series

Assume that the functions $f(x)$ and $g(x)$ are given by the power series

$$f(x) = \sum_{n=0}^{\infty} A_n (x - c)^n \quad g(x) = \sum_{n=0}^{\infty} B_n (x - c)^n$$

for all x obeying $|x - c| < R$. In particular, we are assuming that both power series have radius of convergence at least R . Also let K be a constant. Then

$$\begin{aligned} f(x) + g(x) &= \sum_{n=0}^{\infty} [A_n + B_n] (x - c)^n \\ Kf(x) &= \sum_{n=0}^{\infty} K A_n (x - c)^n \\ (x - c)^N f(x) &= \sum_{n=0}^{\infty} A_n (x - c)^{n+N} \quad \text{for any integer } N \geq 1 \\ &= \sum_{k=N}^{\infty} A_{k-N} (x - c)^k \quad \text{where } k = n + N \\ f'(x) &= \sum_{n=0}^{\infty} A_n n (x - c)^{n-1} = \sum_{n=1}^{\infty} A_n n (x - c)^{n-1} \\ \int_c^x f(t) dt &= \sum_{n=0}^{\infty} A_n \frac{(x - c)^{n+1}}{n+1} \\ \int f(x) dx &= \left[\sum_{n=0}^{\infty} A_n \frac{(x - c)^{n+1}}{n+1} \right] + C \quad \text{with } C \text{ an arbitrary constant} \end{aligned}$$

for all x obeying $|x - c| < R$.

In particular the radius of convergence of each of the six power series on the right hand sides is at least R . In fact, if R is the radius of convergence of $\sum_{n=0}^{\infty} A_n(x-c)^n$, then R is also the radius of convergence of all of the above right hand sides, with the possible exceptions of $\sum_{n=0}^{\infty} [A_n + B_n](x-c)^n$ and $\sum_{n=0}^{\infty} KA_n(x-c)^n$ when $K = 0$.

✓ Example 3.5.14 More on the last part of Theorem 3.5.13

The last statement of Theorem 3.5.13 might seem a little odd, but consider the following two power series centered at 0:

$$\sum_{n=0}^{\infty} 2^n x^n \quad \text{and} \quad \sum_{n=0}^{\infty} (1-2^n)x^n.$$

The ratio test tells us that they both have radius of convergence $R = \frac{1}{2}$. However their sum is

$$\sum_{n=0}^{\infty} 2^n x^n + \sum_{n=0}^{\infty} (1-2^n)x^n = \sum_{n=0}^{\infty} x^n$$

which has the larger radius of convergence 1.

A more extreme example of the same phenomenon is supplied by the two series

$$\sum_{n=0}^{\infty} 2^n x^n \quad \text{and} \quad \sum_{n=0}^{\infty} (-2^n)x^n.$$

They are both geometric series with radius of convergence $R = \frac{1}{2}$. But their sum is

$$\sum_{n=0}^{\infty} 2^n x^n + \sum_{n=0}^{\infty} (-2^n)x^n = \sum_{n=0}^{\infty} (0)x^n$$

which has radius of convergence $+\infty$.

We'll now use this theorem to build power series representations for a bunch of functions out of the one simple power series representation that we know — the geometric series

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \quad \text{for all } |x| < 1$$

✓ Example 3.5.15 $\frac{1}{1-x^2}$

Find a power series representation for $\frac{1}{1-x^2}$.

Solution: The secret to finding power series representations for a good many functions is to manipulate them into a form in which $\frac{1}{1-y}$ appears and use the geometric series representation $\frac{1}{1-y} = \sum_{n=0}^{\infty} y^n$. We have deliberately renamed the variable to y here — it does not have to be x . We can use that strategy to find a power series expansion for $\frac{1}{1-x^2}$ — we just have to recognize that $\frac{1}{1-x^2}$ is the same as $\frac{1}{1-y}$ if we set y to x^2 .

$$\begin{aligned} \frac{1}{1-x^2} &= \frac{1}{1-y} \Big|_{y=x^2} = \left[\sum_{n=0}^{\infty} y^n \right]_{y=x^2} \quad \text{if } |y| < 1, \text{ i.e. } |x| < 1 \\ &= \sum_{n=0}^{\infty} (x^2)^n = \sum_{n=0}^{\infty} x^{2n} \\ &= 1 + x^2 + x^4 + x^6 + \dots \end{aligned}$$

This is a perfectly good power series. There is nothing wrong with the power of x being $2n$. (This just means that the coefficients of all odd powers of x are zero.) In fact, you should try to always write power series in forms that are as easy to understand as possible. The geometric series that we used at the end of the first line converges for

$$|y| < 1 \iff |x^2| < 1 \iff |x| < 1$$

So our power series has radius of convergence 1 and interval of convergence $-1 < x < 1$.

✓ Example 3.5.16 $\frac{x}{2+x^2}$

Find a power series representation for $\frac{x}{2+x^2}$.

Solution: This example is just a more algebraically involved variant of the last one. Again, the strategy is to manipulate $\frac{x}{2+x^2}$ into a form in which $\frac{1}{1-y}$ appears.

$$\begin{aligned} \frac{x}{2+x^2} &= \frac{x}{2} \frac{1}{1+\frac{x^2}{2}} = \frac{x}{2} \frac{1}{1-\left(-\frac{x^2}{2}\right)} && \text{set } -\frac{x^2}{2} = y \\ &= \frac{x}{2} \frac{1}{1-y} \Big|_{y=-\frac{x^2}{2}} = \frac{x}{2} \left[\sum_{n=0}^{\infty} y^n \right]_{y=-\frac{x^2}{2}} && \text{if } |y| < 1 \end{aligned}$$

Now use Theorem 3.5.13 twice

$$\begin{aligned} &= \frac{x}{2} \sum_{n=0}^{\infty} \left(-\frac{x^2}{2}\right)^n = \frac{x}{2} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n} x^{2n} = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^{n+1}} x^{2n+1} \\ &= \frac{x}{2} - \frac{x^3}{4} + \frac{x^5}{8} - \frac{x^7}{16} + \dots \end{aligned}$$

The geometric series that we used in the second line converges when

$$\begin{aligned} |y| < 1 &\iff \left|-\frac{x^2}{2}\right| < 1 \\ &\iff |x|^2 < 2 \iff |x| < \sqrt{2} \end{aligned}$$

So the given power series has radius of convergence $\sqrt{2}$ and interval of convergence $-\sqrt{2} < x < \sqrt{2}$.

✓ Example 3.5.17 Nonzero center

Find a power series representation for $\frac{1}{5-x}$ with center 3.

Solution: The new wrinkle in this example is the requirement that the center be 3. That the center is to be 3 means that we need a power series in powers of $x - c$, with $c = 3$. So we are looking for a power series of the form $\sum_{n=0}^{\infty} A_n (x - 3)^n$. The easy way to find such a series is to force an $x - 3$ to appear by adding and subtracting a 3.

$$\frac{1}{5-x} = \frac{1}{5-(x-3)-3} = \frac{1}{2-(x-3)}$$

Now we continue, as in the last example, by manipulating $\frac{1}{2-(x-3)}$ into a form in which $\frac{1}{1-y}$ appears.

$$\begin{aligned} \frac{1}{5-x} &= \frac{1}{2-(x-3)} = \frac{1}{2} \frac{1}{1-\frac{x-3}{2}} && \text{set } \frac{x-3}{2} = y \\ &= \frac{1}{2} \frac{1}{1-y} \Big|_{y=\frac{x-3}{2}} = \frac{1}{2} \left[\sum_{n=0}^{\infty} y^n \right]_{y=\frac{x-3}{2}} && \text{if } |y| < 1 \\ &= \frac{1}{2} \sum_{n=0}^{\infty} \left(\frac{x-3}{2}\right)^n = \sum_{n=0}^{\infty} \frac{(x-3)^n}{2^{n+1}} \\ &= \frac{x-3}{2} + \frac{(x-3)^2}{4} + \frac{(x-3)^3}{8} + \dots \end{aligned}$$

The geometric series that we used in the second line converges when

$$\begin{aligned}
 |y| < 1 &\iff \left| \frac{x-3}{2} \right| < 1 \\
 &\iff |x-3| < 2 \\
 &\iff -2 < x-3 < 2 \\
 &\iff 1 < x < 5
 \end{aligned}$$

So the power series has radius of convergence 2 and interval of convergence $1 < x < 5$.

In the previous two examples, to construct a new series from an existing series, we replaced x by a simple function. The following theorem gives us some more (but certainly not all) commonly used substitutions.

Theorem 3.5.18 Substituting in a Power Series.

Assume that the function $f(x)$ is given by the power series

$$f(x) = \sum_{n=0}^{\infty} A_n x^n$$

for all x in the interval I . Also let K and k be real constants. Then

$$f(Kx^k) = \sum_{n=0}^{\infty} A_n K^n x^{kn}$$

whenever Kx^k is in I . In particular, if $\sum_{n=0}^{\infty} A_n x^n$ has radius of convergence R , K is nonzero and k is a natural number, then $\sum_{n=0}^{\infty} A_n K^n x^{kn}$ has radius of convergence $\sqrt[k]{R/|K|}$.

Example 3.5.19 $\frac{1}{(1-x)^2}$

Find a power series representation for $\frac{1}{(1-x)^2}$.

Solution: Once again the trick is to express $\frac{1}{(1-x)^2}$ in terms of $\frac{1}{1-x}$. Notice that

$$\begin{aligned}
 \frac{1}{(1-x)^2} &= \frac{d}{dx} \left\{ \frac{1}{1-x} \right\} \\
 &= \frac{d}{dx} \left\{ \sum_{n=0}^{\infty} x^n \right\} \\
 &= \sum_{n=1}^{\infty} n x^{n-1} \quad \text{by Theorem 3.5.13}
 \end{aligned}$$

Note that the $n = 0$ term has disappeared because, for $n = 0$,

$$\frac{d}{dx} x^n = \frac{d}{dx} x^0 = \frac{d}{dx} 1 = 0$$

Also note that the radius of convergence of this series is one. We can see this via Theorem 3.5.13. That theorem tells us that the radius of convergence of a power series is not changed by differentiation — and since $\sum_{n=0}^{\infty} x^n$ has radius of convergence one, so too does its derivative.

Without much more work we can determine the interval of convergence by testing at $x = \pm 1$. When $x = \pm 1$ the terms of the series do not go to zero as $n \rightarrow \infty$ and so, by the divergence test, the series does not converge there. Hence the interval of convergence for the series is $-1 < x < 1$.

Notice that, in this last example, we differentiated a known series to get to our answer. As per Theorem 3.5.13, the radius of convergence didn't change. In addition, in this particular example, the interval of convergence didn't change. This is not always the case. Differentiation of some series causes the interval of convergence to shrink. In particular the differentiated series may no

longer be convergent at the end points of the interval⁸ Similarly, when we integrate a power series the radius of convergence is unchanged, but the interval of convergence may expand to include one or both ends, as illustrated by the next example.

✓ Example 3.5.20 $\log(1+x)$

Find a power series representation for $\log(1+x)$.

Solution: Recall that $\frac{d}{dx} \log(1+x) = \frac{1}{1+x}$ so that $\log(1+t)$ is an antiderivative of $\frac{1}{1+t}$ and

$$\begin{aligned} \log(1+x) &= \int_0^x \frac{dt}{1+t} = \int_0^x \left[\sum_{n=0}^{\infty} (-t)^n \right] dt \\ &= \sum_{n=0}^{\infty} \int_0^x (-t)^n dt \quad \text{by Theorem 3.5.13} \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} \\ &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \end{aligned}$$

Theorem 3.5.13 guarantees that the radius of convergence is exactly one (the radius of convergence of the geometric series $\sum_{n=0}^{\infty} (-t)^n$) and that

$$\log(1+x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} \quad \text{for all } -1 < x < 1$$

When $x = -1$ our series reduces to $\sum_{n=0}^{\infty} \frac{-1}{n+1}$, which is (minus) the harmonic series and so diverges. That's no surprise — $\log(1+(-1)) = \log 0 = -\infty$. When $x = 1$, the series converges by the alternating series test. It is possible to prove, by continuity, though we won't do so here, that the sum is $\log 2$. So the interval of convergence is $-1 < x \leq 1$.

✓ Example 3.5.21 $\arctan x$

Find a power series representation for $\arctan x$.

Solution: Recall that $\frac{d}{dx} \arctan x = \frac{1}{1+x^2}$ so that $\arctan t$ is an antiderivative of $\frac{1}{1+t^2}$ and

$$\begin{aligned} \arctan x &= \int_0^x \frac{dt}{1+t^2} = \int_0^x \left[\sum_{n=0}^{\infty} (-t^2)^n \right] dt = \sum_{n=0}^{\infty} \int_0^x (-1)^n t^{2n} dt \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} \\ &= x - \frac{x^3}{3} + \frac{x^5}{5} - \dots \end{aligned}$$

Theorem 3.5.13 guarantees that the radius of convergence is exactly one (the radius of convergence of the geometric series $\sum_{n=0}^{\infty} (-t^2)^n$) and that

$$\arctan x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} \quad \text{for all } -1 < x < 1$$

When $x = \pm 1$, the series converges by the alternating series test. So the interval of convergence is $-1 \leq x \leq 1$. It is possible to prove, though once again we won't do so here, that when $x = \pm 1$, the series $\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$ converges to the value of the left hand side, $\arctan x$, at $x = \pm 1$. That is, to $\arctan(\pm 1) = \pm \frac{\pi}{4}$.

The operations on power series dealt with in Theorem 3.5.13 are fairly easy to apply. Unfortunately taking the product, ratio or composition of two power series is more involved and is beyond the scope of this course⁹. Unfortunately Theorem 3.5.13 alone will not get us power series representations of many of our standard functions (like e^x and $\sin x$). Fortunately we can find such representations by extending Taylor polynomials¹⁰ to Taylor series.

Exercises

Stage 1

? 1

Suppose $f(x) = \sum_{n=0}^{\infty} \left(\frac{3-x}{4}\right)^n$. What is $f(1)$?

? 2

Suppose $f(x) = \sum_{n=1}^{\infty} \frac{(x-5)^n}{n!+2}$. Give a power series representation of $f'(x)$.

? 3

Let $f(x) = \sum_{n=a}^{\infty} A_n(x-c)^n$ for some positive constants a and c , and some sequence of constants $\{A_n\}$. For which values of x does $f(x)$ definitely converge?

? 4

Let $f(x)$ be a power series centered at $c = 5$. If $f(x)$ converges at $x = -1$, and diverges at $x = 11$, what is the radius of convergence of $f(x)$?

Stage 2

? 5 (*)

(a) Find the radius of convergence of the series

$$\sum_{k=0}^{\infty} (-1)^k 2^{k+1} x^k$$

(b) You are given the formula for the sum of a geometric series, namely:

$$1 + r + r^2 + \dots = \frac{1}{1-r}, \quad |r| < 1$$

Use this fact to evaluate the series in part (a).

? 6 (*)

Find the radius of convergence for the power series $\sum_{k=0}^{\infty} \frac{x^k}{10^{k+1}(k+1)!}$

? 7 (*)

Find the radius of convergence for the power series $\sum_{n=0}^{\infty} \frac{(x-2)^n}{n^2+1}$.

? 8 (*)

Consider the power series $\sum_{n=1}^{\infty} \frac{(-1)^n (x+2)^n}{\sqrt{n}}$, where x is a real number. Find the interval of convergence of this series.

? 9 (*)

Find the radius of convergence and interval of convergence of the series

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} \left(\frac{x+1}{3} \right)^n$$

? 10 (*)

Find the *interval of convergence* for the power series

$$\sum_{n=1}^{\infty} \frac{(x-2)^n}{n^{4/5}(5^n - 4)}$$

? 11 (*)

Find all values x for which the series $\sum_{n=1}^{\infty} \frac{(x+2)^n}{n^2}$ converges.

? 12 (*)

Find the interval of convergence for $\sum_{n=1}^{\infty} \frac{4^n}{n} (x-1)^n$.

? 13 (*)

Find, with explanation, the radius of convergence and the interval of convergence of the power series

$$\sum_{n=0}^{\infty} (-1)^n \frac{(x-1)^n}{2^n(n+2)}$$

? 14 (*)

Find the interval of convergence for the series $\sum_{n=1}^{\infty} (-1)^n n^2 (x-a)^{2n}$ where a is a constant.

? 15 (*)

Find the interval of convergence of the following series:

1. $\sum_{k=1}^{\infty} \frac{(x+1)^k}{k^2 9^k}$.

2. $\sum_{k=1}^{\infty} a_k (x-1)^k$, where $a_k > 0$ for $k = 1, 2, \dots$ and $\sum_{k=1}^{\infty} \left(\frac{a_k}{a_{k+1}} - \frac{a_{k+1}}{a_{k+2}} \right) = \frac{a_1}{a_2}$.

? 16 (*)

Find a power series representation for $\frac{x^3}{1-x}$.

? 17

Suppose $f'(x) = \sum_{n=0}^{\infty} \frac{(x-1)^n}{n+2}$, and $\int_5^x f(t) dt = 3x + \sum_{n=1}^{\infty} \frac{(x-1)^{n+1}}{n(n+1)^2}$.

Give a power series representation of $f(x)$.

Stage 3

? 18 (*)

Determine the values of x for which the series

$$\sum_{n=2}^{\infty} \frac{x^n}{3^{2n} \log n}$$

converges absolutely, converges conditionally, or diverges.

? 19 (*)

(a) Find the power-series representation for $\int \frac{1}{1+x^3} dx$ centered at 0 (i.e. in powers of x).

(b) The power series above is used to approximate $\int_0^{1/4} \frac{1}{1+x^3} dx$. How many terms are required to guarantee that the resulting approximation is within 10^{-5} of the exact value? Justify your answer.

? 20 (*)

(a) Show that $\sum_{n=0}^{\infty} nx^n = \frac{x}{(1-x)^2}$ for $-1 < x < 1$.

(b) Express $\sum_{n=0}^{\infty} n^2 x^n$ as a ratio of polynomials. For which x does this series converge?

? 21 (*)

Suppose that you have a sequence $\{b_n\}$ such that the series $\sum_{n=0}^{\infty} (1-b_n)$ converges. Using the tests we've learned in class, prove that the radius of convergence of the power series $\sum_{n=0}^{\infty} b_n x^n$ is equal to 1.

? 22 (*)

Assume $\{a_n\}$ is a sequence such that na_n decreases to C as $n \rightarrow \infty$ for some real number $C > 0$

(a) Find the radius of convergence of $\sum_{n=1}^{\infty} a_n x^n$. Justify your answer carefully.

(b) Find the interval of convergence of the above power series, that is, find all x for which the power series in (a) converges. Justify your answer carefully.

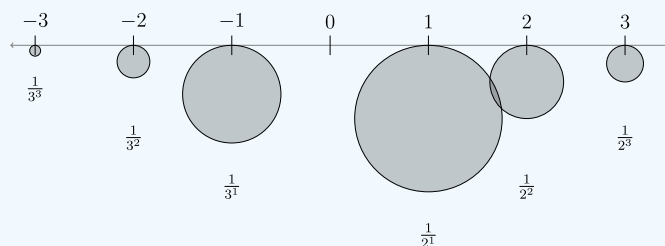
? 23

An infinitely long, straight rod of negligible mass has the following weights:

- At every whole number n , a mass of weight $\frac{1}{2^n}$ at position n , and

- a mass of weight $\frac{1}{3^n}$ at position $-n$.

At what position is the center of mass of the rod?



? 24

Let $f(x) = \sum_{n=0}^{\infty} A_n(x-c)^n$, for some constant c and a sequence of constants $\{A_n\}$. Further, let $f(x)$ have a positive radius of convergence.

If $A_1 = 0$, show that $y = f(x)$ has a critical point at $x = c$. What is the relationship between the behaviour of the graph at that point and the value of A_2 ?

? 25

Evaluate $\sum_{n=3}^{\infty} \frac{n}{5^{n-1}}$.

? 26

Find a polynomial that approximates $f(x) = \log(1+x)$ to within an error of 10^{-5} for all values of x in $(0, \frac{1}{10})$.

Then, use your polynomial to approximate $\log(1.05)$ as a rational number.

? 27

Find a polynomial that approximates $f(x) = \arctan x$ to within an error of 10^{-5} for all values of x in $(-\frac{1}{4}, \frac{1}{4})$.

1. Recall that $n! = 1 \times 2 \times 3 \times \dots \times n$ is called “ n factorial”. By convention $0! = 1$.
2. *By convention*, when the term $(x-c)^0$ appears in a power series, it has value 1 for all values of x , even $x = c$.
3. The use of the word “radius” might seem a little odd here, since we are really describing the interval in the real line where the series converges. However, when one starts to consider power series over complex numbers, the radius of convergence does describe a circle inside the complex plane and so “radius” is a more natural descriptor.
4. Because of this, it might seem that such a series is fairly pointless. However there are all sorts of mathematical games that can be played with them without worrying about their convergence. Such “formal” power series can still impart useful information and the interested reader is invited to look up “generating functions” with their preferred search engine.
5. We give a proof of this in the optional §3.7 at the end of this chapter.
6. This is a little beyond the scope of the course. Roughly speaking, think about what would happen if the limit of the ratios did exist. If the limit were smaller than 1, then it would tell you that the terms of our series must be getting smaller and smaller and smaller — which is impossible because they are all integers between 1 and 10. Similarly if the limit existed and were bigger than 1 then the terms of the series would have to get bigger and bigger and bigger — also impossible. Hence if the ratio exists then it must be equal to 1 — but in that case because the terms are integers, they would have to be all equal when n became big enough. But that means that the expansion of π would be eventually periodic — something that only rational numbers do (a proof is given in the optional §3.7 at the end of this chapter).
7. We avoid problems with 0^0 by starting the series from $n = 1$.

8. Consider the power series $\sum_{n=1}^{\infty} \frac{x^n}{n}$. We know that its interval of convergence is $-1 \leq x < 1$. (Indeed see the next example.)
When we differentiate the series we get the geometric series $\sum_{n=0}^{\infty} x^n$ which has interval of convergence $-1 < x < 1$.
 9. As always, a quick visit to your favourite search engine will direct the interested reader to more information.
 10. Now is a good time to review your notes from last term, though we'll give you a whirlwind review over the next page or two.
-

This page titled [3.5: Power Series](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

3.6: Taylor Series

Extending Taylor Polynomials

Recall¹ that Taylor polynomials provide a hierarchy of approximations to a given function $f(x)$ near a given point a . Typically, the quality of these approximations improves as we move up the hierarchy.

- The crudest approximation is the constant approximation $f(x) \approx f(a)$.
- Then comes the linear, or tangent line, approximation $f(x) \approx f(a) + f'(a)(x - a)$.
- Then comes the quadratic approximation

$$f(x) \approx f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2$$

- In general, the Taylor polynomial of degree n , for the function $f(x)$, about the expansion point a , is the polynomial, $T_n(x)$, determined by the requirements that $f^{(k)}(a) = T_n^{(k)}(a)$ for all $0 \leq k \leq n$. That is, f and T_n have the same derivatives at a , up to order n . Explicitly,

$$\begin{aligned} f(x) &\approx T_n(x) \\ &= f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2 + \cdots + \frac{1}{n!}f^{(n)}(a)(x - a)^n \\ &= \sum_{k=0}^n \frac{1}{k!}f^{(k)}(a)(x - a)^k \end{aligned}$$

These are, of course, approximations — often very good approximations near $x = a$ — but still just approximations. One might hope that if we let the degree, n , of the approximation go to infinity then the error in the approximation might go to zero. If that is the case then the “infinite” Taylor polynomial would be an exact representation of the function. Let's see how this might work.

Fix a real number a and suppose that all derivatives of the function $f(x)$ exist. Then, we saw in (3.4.33) of the CLP-1 text that, for any natural number n ,

Equation 3.6.1

$$f(x) = T_n(x) + E_n(x)$$

where $T_n(x)$ is the Taylor polynomial of degree n for the function $f(x)$ expanded about a , and $E_n(x) = f(x) - T_n(x)$ is the error in the approximation $f(x) \approx T_n(x)$. The Taylor polynomial² is given by the formula

Equation 3.6.2

$$T_n(x) = f(a) + f'(a)(x - a) + \cdots + \frac{1}{n!}f^{(n)}(a)(x - a)^n$$

while the error satisfies³

Equation 3.6.3

$$E_n(x) = \frac{1}{(n+1)!}f^{(n+1)}(c)(x - a)^{n+1}$$

for some c strictly between a and x .

Note that we typically do not know the value of c in the formula for the error. Instead we use the bounds on c to find bounds on $f^{(n+1)}(c)$ and so bound the error⁴.

In order for our Taylor polynomial to be an exact representation of the function $f(x)$ we need the error $E_n(x)$ to be zero. This will not happen when n is finite unless $f(x)$ is a polynomial. However it can happen in the limit as $n \rightarrow \infty$, and in that case we can write $f(x)$ as the limit

$$f(x) = \lim_{n \rightarrow \infty} T_n(x) = \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{1}{k!} f^{(k)}(a) (x-a)^k$$

This is really a limit of partial sums, and so we can write

$$f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(a) (x-a)^k$$

which is a power series representation of the function. Let us formalise this in a definition.

Definition 3.6.4 Taylor series

The Taylor series for the function $f(x)$ expanded around a is the power series

$$\sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(a) (x-a)^n$$

When $a = 0$ it is also called the Maclaurin series of $f(x)$. If $\lim_{n \rightarrow \infty} E_n(x) = 0$, then

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(a) (x-a)^n$$

Demonstrating that, for a given function, $\lim_{n \rightarrow \infty} E_n(x) = 0$ can be difficult, but for many of the standard functions you are used to dealing with, it turns out to be pretty easy. Let's compute a few Taylor series and see how we do it.

Example 3.6.5 Exponential Series

Find the Maclaurin series for $f(x) = e^x$.

Solution

Just as was the case for computing Taylor polynomials, we need to compute the derivatives of the function at the particular choice of a . Since we are asked for a Maclaurin series, $a = 0$. So now we just need to find $f^{(k)}(0)$ for all integers $k \geq 0$.

We know that $\frac{d}{dx} e^x = e^x$ and so

$$\begin{aligned} e^x &= f(x) = f'(x) = f''(x) = \dots = f^{(k)}(x) = \dots && \text{which gives} \\ 1 &= f(0) = f'(0) = f''(0) = \dots = f^{(k)}(0) = \dots \end{aligned}$$

Equations 3.6.1 and 3.6.2 then give us

$$e^x = f(x) = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + E_n(x)$$

We shall see, in the optional Example 3.6.8 below, that, for any fixed x , $\lim_{n \rightarrow \infty} E_n(x) = 0$. Consequently, for all x ,

$$e^x = \lim_{n \rightarrow \infty} \left[1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots + \frac{1}{n!}x^n \right] = \sum_{n=0}^{\infty} \frac{1}{n!}x^n$$

We have now seen power series representations for the functions

$$\frac{1}{1-x} \quad \frac{1}{(1-x)^2} \quad \log(1+x) \quad \arctan(x) \quad e^x.$$

We do not think that you, the reader, will be terribly surprised to see that we develop series for sine and cosine next.

✓ Example 3.6.6 Sine and Cosine Series

The trigonometric functions $\sin x$ and $\cos x$ also have widely used Maclaurin series expansions (i.e. Taylor series expansions about $a = 0$). To find them, we first compute all derivatives at general x .

$$\begin{array}{llll} f(x) = \sin x & f'(x) = \cos x & f''(x) = -\sin x & f^{(3)}(x) = -\cos x \\ & f^{(4)}(x) = \sin x & \dots & \\ g(x) = \cos x & g'(x) = -\sin x & g''(x) = -\cos x & g^{(3)}(x) = \sin x \\ & g^{(4)}(x) = \cos x & \dots & \end{array}$$

Now set $x = a = 0$.

$$\begin{array}{llllll} f(x) = \sin x & f(0) = 0 & f'(0) = 1 & f''(0) = 0 & f^{(3)}(0) = -1 \\ & f^{(4)}(0) = 0 & \dots & & \\ g(x) = \cos x & g(0) = 1 & g'(0) = 0 & g''(0) = -1 & g^{(3)}(0) = 0 \\ & g^{(4)}(0) = 1 & \dots & & \end{array}$$

For $\sin x$, all even numbered derivatives (at $x = 0$) are zero, while the odd numbered derivatives alternate between 1 and -1 . Very similarly, for $\cos x$, all odd numbered derivatives (at $x = 0$) are zero, while the even numbered derivatives alternate between 1 and -1 . So, the Taylor polynomials that best approximate $\sin x$ and $\cos x$ near $x = a = 0$ are

$$\begin{aligned} \sin x &\approx x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots \\ \cos x &\approx 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \dots \end{aligned}$$

We shall see, in the optional Example 3.6.10 below, that, for both $\sin x$ and $\cos x$, we have $\lim_{n \rightarrow \infty} E_n(x) = 0$ so that

$$\begin{aligned} f(x) &= \lim_{n \rightarrow \infty} \left[f(0) + f'(0)x + \dots + \frac{1}{n!}f^{(n)}(0)x^n \right] \\ g(x) &= \lim_{n \rightarrow \infty} \left[g(0) + g'(0)x + \dots + \frac{1}{n!}g^{(n)}(0)x^n \right] \end{aligned}$$

Reviewing the patterns we found in the derivatives, we conclude that, for all x ,

$$\begin{aligned} \sin x &= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots = \sum_{n=0}^{\infty} (-1)^n \frac{1}{(2n+1)!} x^{2n+1} \\ \cos x &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \dots = \sum_{n=0}^{\infty} (-1)^n \frac{1}{(2n)!} x^{2n} \end{aligned}$$

and, in particular, both of the series on the right hand sides converge for all x .

We could also test for convergence of the series using the ratio test. Computing the ratios of successive terms in these two series gives us

$$\begin{aligned} \left| \frac{A_{n+1}}{A_n} \right| &= \frac{|x|^{2n+3}/(2n+3)!}{|x|^{2n+1}/(2n+1)!} = \frac{|x|^2}{(2n+3)(2n+2)} \\ \left| \frac{A_{n+1}}{A_n} \right| &= \frac{|x|^{2n+2}/(2n+2)!}{|x|^{2n}/(2n)!} = \frac{|x|^2}{(2n+2)(2n+1)} \end{aligned}$$

for sine and cosine respectively. Hence as $n \rightarrow \infty$ these ratios go to zero and consequently both series are convergent for all x . (This is very similar to what was observed in Example 3.5.5.)

We have developed power series representations for a number of important functions⁵. Here is a theorem that summarizes them.

Theorem 3.6.7

$$\begin{aligned}
 e^x &= \sum_{n=0}^{\infty} \frac{x^n}{n!} &&= 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots && \text{for all } -\infty < x < \infty \\
 \sin(x) &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} &&= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots && \text{for all } -\infty < x < \infty \\
 \cos(x) &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} &&= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \dots && \text{for all } -\infty < x < \infty \\
 \frac{1}{1-x} &= \sum_{n=0}^{\infty} x^n &&= 1 + x + x^2 + x^3 + \dots && \text{for all } -1 < x < 1 \\
 \log(1+x) &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} &&= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots && \text{for all } -1 < x \leq 1 \\
 \arctan x &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} &&= x - \frac{x^3}{3} + \frac{x^5}{5} - \dots && \text{for all } -1 \leq x \leq 1
 \end{aligned}$$

Notice that the series for sine and cosine sum to something that looks very similar to the series for e^x :

$$\begin{aligned}
 \sin(x) + \cos(x) &= \left(x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots \right) + \left(1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \dots \right) \\
 &= 1 + x - \frac{1}{2!}x^2 - \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \frac{1}{5!}x^5 - \dots \\
 e^x &= 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \frac{1}{5!}x^5 + \dots
 \end{aligned}$$

So both series have coefficients with the same absolute value (namely $\frac{1}{n!}$), but there are differences in sign⁶. This is not a coincidence and we direct the interested reader to the optional Section 3.6.3 where will show how these series are linked through $\sqrt{-1}$.

✓ Example 3.6.8 Optional — Why $\sum_{n=0}^{\infty} \frac{1}{n!}x^n$ is e^x

We have already seen, in Example 3.6.5, that

$$e^x = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + E_n(x)$$

By (3.6.3)

$$E_n(x) = \frac{1}{(n+1)!} e^c x^{n+1}$$

for some (unknown) c between 0 and x . Fix any real number x . We'll now show that $E_n(x)$ converges to zero as $n \rightarrow \infty$.

To do this we need get bound the size of e^c , and to do this, consider what happens if x is positive or negative.

- If $x < 0$ then $x \leq c \leq 0$ and hence $e^x \leq e^c \leq e^0 = 1$.
- On the other hand, if $x \geq 0$ then $0 \leq c \leq x$ and so $1 = e^0 \leq e^c \leq e^x$.

In either case we have that $0 \leq e^c \leq 1 + e^x$. Because of this the error term

$$|E_n(x)| = \left| \frac{e^c}{(n+1)!} x^{n+1} \right| \leq [e^x + 1] \frac{|x|^{n+1}}{(n+1)!}$$

We claim that this upper bound, and hence the error $E_n(x)$, quickly shrinks to zero as $n \rightarrow \infty$.

Call the upper bound (except for the factor $e^x + 1$, which is independent of n) $e_n(x) = \frac{|x|^{n+1}}{(n+1)!}$. To show that this shrinks to zero as $n \rightarrow \infty$, let's write it as follows.

$$e_n(x) = \frac{|x|^{n+1}}{(n+1)!} = \overbrace{\frac{|x|}{1} \cdot \frac{|x|}{2} \cdot \frac{|x|}{3} \cdots \frac{|x|}{n} \cdot \frac{|x|}{n+1}}^{n+1 \text{ factors}}$$

Now let k be an integer bigger than $|x|$. We can split the product

$$\begin{aligned} e_n(x) &= \left(\frac{|x|}{1} \cdot \frac{|x|}{2} \cdot \frac{|x|}{3} \cdots \frac{|x|}{k} \right) \cdot \left(\frac{|x|}{k+1} \cdots \frac{|x|}{n+1} \right) \\ &\leq \underbrace{\left(\frac{|x|}{1} \cdot \frac{|x|}{2} \cdot \frac{|x|}{3} \cdots \frac{|x|}{k} \right)}_{=Q(x)} \cdot \left(\frac{|x|}{k+1} \right)^{n+1-k} \\ &= Q(x) \cdot \left(\frac{|x|}{k+1} \right)^{n+1-k} \end{aligned}$$

Since k does not depend on n (though it does depend on x), the function $Q(x)$ does not change as we increase n . Additionally, we know that $|x| < k+1$ and so $\frac{|x|}{k+1} < 1$. Hence as we let $n \rightarrow \infty$ the above bound must go to zero.

Alternatively, compare $e_n(x)$ and $e_{n+1}(x)$.

$$\frac{e_{n+1}(x)}{e_n(x)} = \frac{\frac{|x|^{n+2}}{(n+2)!}}{\frac{|x|^{n+1}}{(n+1)!}} = \frac{|x|}{n+2}$$

When n is bigger than, for example $2|x|$, we have $\frac{e_{n+1}(x)}{e_n(x)} < \frac{1}{2}$. That is, increasing the index on $e_n(x)$ by one decreases the size of $e_n(x)$ by a factor of at least two. As a result $e_n(x)$ must tend to zero as $n \rightarrow \infty$.

Consequently, for all x , $\lim_{n \rightarrow \infty} E_n(x) = 0$, as claimed, and we really have

$$e^x = \lim_{n \rightarrow \infty} \left[1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n \right] = \sum_{n=0}^{\infty} \frac{1}{n!}x^n$$

There is another way to prove that the series $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges to the function e^x . Rather than looking at how the error term $E_n(x)$ behaves as $n \rightarrow \infty$, we can show that the series satisfies the same simple differential equation⁷ and the same initial condition as the function.

✓ **Example 3.6.9 Optional** — Another approach to showing that $\sum_{n=0}^{\infty} \frac{1}{n!}x^n$ is e^x

We already know from Example 3.5.5, that the series $\sum_{n=0}^{\infty} \frac{1}{n!}x^n$ converges to some function $f(x)$ for all values of x . All that remains to do is to show that $f(x)$ is really e^x . We will do this by showing that $f(x)$ and e^x satisfy the same differential equation with the same initial conditions⁸. We know that $y = e^x$ satisfies

$$\frac{dy}{dx} = y \quad \text{and} \quad y(0) = 1$$

and by Theorem 2.4.4 (with $a = 1$, $b = 0$ and $y(0) = 1$), this is the only solution. So it suffices to show that $f(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ satisfies

$$\frac{df}{dx} = f(x) \quad \text{and} \quad f(0) = 1.$$

- By Theorem 3.5.13,

$$\begin{aligned} \frac{df}{dx} &= \frac{d}{dx} \left\{ \sum_{n=0}^{\infty} \frac{1}{n!} x^n \right\} = \sum_{n=1}^{\infty} \frac{n}{n!} x^{n-1} = \sum_{n=1}^{\infty} \frac{1}{(n-1)!} x^{n-1} \\ &= \underbrace{\frac{1}{1}}_{n=1} + \underbrace{\frac{x}{1}}_{n=2} + \underbrace{\frac{x^2}{2!}}_{n=3} + \underbrace{\frac{x^3}{3!}}_{n=4} + \dots \\ &= f(x) \end{aligned}$$

- When we substitute $x = 0$ into the series we get (see the discussion after Definition 3.5.1)

$$f(0) = 1 + \frac{0}{1!} + \frac{0}{2!} + \dots = 1.$$

Hence $f(x)$ solves the same initial value problem and we must have $f(x) = e^x$.

We can show that the error terms in Maclaurin polynomials for sine and cosine go to zero as $n \rightarrow \infty$ using very much the same approach as in Example 3.6.8.

✓ Example 3.6.10 Optional — Why $\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = \sin x$ and $\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = \cos x$

Let $f(x)$ be either $\sin x$ or $\cos x$. We know that every derivative of $f(x)$ will be one of $\pm \sin(x)$ or $\pm \cos(x)$. Consequently, when we compute the error term using equation 3.6.3 we always have $|f^{(n+1)}(c)| \leq 1$ and hence

$$|E_n(x)| \leq \frac{|x|^{n+1}}{(n+1)!}.$$

In Example 3.6.5, we showed that $\frac{|x|^{n+1}}{(n+1)!} \rightarrow 0$ as $n \rightarrow \infty$ — so all the hard work is already done. Since the error term shrinks to zero for both $f(x) = \sin x$ and $f(x) = \cos x$, and

$$f(x) = \lim_{n \rightarrow \infty} \left[f(0) + f'(0)x + \dots + \frac{1}{n!} f^{(n)}(0) x^n \right]$$

as required.

Optional — More about the Taylor Remainder

In this section, we fix a real number a and a natural number n , suppose that all derivatives of the function $f(x)$ exist, and we study the error

$$\begin{aligned} E_n(a, x) &= f(x) - T_n(a, x) \\ \text{where } T_n(a, x) &= f(a) + f'(a)(x-a) + \dots + \frac{1}{n!} f^{(n)}(a)(x-a)^n \end{aligned}$$

made when we approximate $f(x)$ by the Taylor polynomial $T_n(a, x)$ of degree n for the function $f(x)$, expanded about a . We have already seen, in (3.6.3), one formula, probably the most commonly used formula, for $E_n(a, x)$. In the next theorem, we repeat that formula and give a second, commonly used, formula. After an example, we give a second theorem that contains some less commonly used formulae.

🔗 Theorem 3.6.11 Commonly used formulae for the Taylor remainder

The Taylor remainder $E_n(a, x)$ is given by

1. (integral form)

$$E_n(a, x) = \int_a^x \frac{1}{n!} f^{(n+1)}(t) (x-t)^n dt$$

2. (Lagrange form)

$$E_n(a, x) = \frac{1}{(n+1)!} f^{(n+1)}(c) (x-a)^{n+1}$$

for some c strictly between a and x .

Notice that the integral form of the error is explicit - we could, in principle, compute it exactly. (Of course if we could do that, we probably wouldn't need to use a Taylor expansion to approximate f .) This contrasts with the Lagrange form which is an 'existential' statement - it tells us that ' c ' exists, but not how to compute it.

Proof

1. We will give two proofs. The first is shorter and simpler, but uses some trickery. The second is longer, but is more straightforward. It uses a technique called mathematical induction.

Proof 1: We are going to use a little trickery to get a simple proof. We simply view x as being fixed and study the dependence of $E_n(a, x)$ on a . To emphasise that that is what we are doing, we define

$$S(t) = f(x) - f(t) - f'(t)(x-t) - \frac{1}{2}f''(t)(x-t)^2 - \dots - \frac{1}{n!}f^{(n)}(t)(x-t)^n \quad (*)$$

and observe that $E_n(a, x) = S(a)$.

So, by the fundamental theorem of calculus (Theorem 1.3.1), the function $S(t)$ is determined by its derivative, $S'(t)$, and its value at a single point. Finding a value of $S(t)$ for one value of t is easy. Substitute $t = x$ into (*) to yield $S(x) = 0$. To find $S'(t)$, apply $\frac{d}{dt}$ to both sides of (*). Recalling that x is just a constant parameter,

$$\begin{aligned} S'(t) &= 0 - f'(t) - [f''(t)(x-t) - f'(t)] \\ &\quad - [\frac{1}{2}f^{(3)}(t)(x-t)^2 - f''(t)(x-t)] \\ &\quad - \dots - [\frac{1}{n!}f^{(n+1)}(t)(x-t)^n - \frac{1}{(n-1)!}f^{(n)}(t)(x-t)^{n-1}] \\ &= -\frac{1}{n!}f^{(n+1)}(t)(x-t)^n \end{aligned}$$

So, by the fundamental theorem of calculus, $S(x) = S(a) + \int_a^x S'(t) dt$ and

$$\begin{aligned} E_n(a, x) &= -[S(x) - S(a)] = -\int_a^x S'(t) dt \\ &= \int_a^x \frac{1}{n!}f^{(n+1)}(t)(x-t)^n dt \end{aligned}$$

Proof 2: The proof that we have just given was short, but also very tricky --- almost noone could create that proof without big hints. Here is another much less tricky, but also commonly used, proof.

- o First consider the case $n = 0$. When $n = 0$,

$$E_0(a, x) = f(x) - T_0(a, x) = f(x) - f(a)$$

The fundamental theorem of calculus gives

$$f(x) - f(a) = \int_a^x f'(t) dt$$

so that

$$E_0(a, x) = \int_a^x f'(t) dt$$

That is exactly the $n = 0$ case of part (a).

- o Next fix any integer $n \geq 0$ and suppose that we already know that

$$E_n(a, x) = \int_a^x \frac{1}{n!}f^{(n+1)}(t)(x-t)^n dt$$

Apply integration by parts (Theorem 1.7.2) to this integral with

$$u(t) = f^{(n+1)}(t)$$

$$dv = \frac{1}{n!}(x-t)^n dt, \quad v(t) = -\frac{1}{(n+1)!}(x-t)^{n+1}$$

Since $v(x) = 0$, integration by parts gives

$$\begin{aligned} E_n(a, x) &= u(x)v(x) - u(a)v(a) - \int_a^x v(t)u'(t) dt \\ &= \frac{1}{(n+1)!} f^{(n+1)}(a) (x-a)^{n+1} \\ &\quad + \int_a^x \frac{1}{(n+1)!} f^{(n+2)}(t) (x-t)^{n+1} dt \end{aligned} \quad (**)$$

Now, we defined

$$\begin{aligned} E_n(a, x) &= f(x) - f(a) - f'(a)(x-a) - \frac{1}{2}f''(a)(x-a)^2 \\ &\quad - \dots - \frac{1}{n!}f^{(n)}(a)(x-a)^n \end{aligned}$$

so

$$E_{n+1}(a, x) = E_n(a, x) - \frac{1}{(n+1)!}f^{(n+1)}(a)(x-a)^{n+1}$$

This formula expresses $E_{n+1}(a, x)$ in terms of $E_n(a, x)$. That's called a reduction formula. Combining the reduction formula with (**) gives

$$E_{n+1}(a, x) = \int_a^x \frac{1}{(n+1)!} f^{(n+2)}(t) (x-t)^{n+1} dt$$

- o Let's pause to summarise what we have learned in the last two bullets. Use the notation $P(n)$ to stand for the statement " $E_n(a, x) = \int_a^x \frac{1}{n!} f^{(n+1)}(t) (x-t)^n dt$ ". To prove part (a) of the theorem, we need to prove that the statement $P(n)$ is true for all integers $n \geq 0$. In the first bullet, we showed that the statement $P(0)$ is true. In the second bullet, we showed that if, for some integer $n \geq 0$, the statement $P(n)$ is true, then the statement $P(n+1)$ is also true. Consequently,
 - $P(0)$ is true by the first bullet and then
 - $P(1)$ is true by the second bullet with $n = 0$ and then
 - $P(2)$ is true by the second bullet with $n = 1$ and then
 - $P(3)$ is true by the second bullet with $n = 2$
 - and so on, for ever and ever.

That tells us that $P(n)$ is true for all integers $n \geq 0$, which is exactly part (a) of the theorem. This proof technique is called mathematical induction⁹.

2. We have already seen one proof in the optional Section 3.4.9 of the CLP-1 text. We will see two more proofs here.

Proof 1: We apply the generalised mean value theorem, which is Theorem 3.4.38 in the CLP-1 text. It says that

$$\frac{F(b) - F(a)}{G(b) - G(a)} = \frac{F'(c)}{G'(c)}$$

for some c strictly between¹⁰ a and b . We apply (GMVT) with $b = x$, $F(t) = S(t)$ and $G(t) = (x-t)^{n+1}$. This gives

$$\begin{aligned} E_n(a, x) &= -[S(x) - S(a)] = -\frac{S'(c)}{G'(c)} [G(x) - G(a)] \\ &= -\frac{-\frac{1}{n!}f^{(n+1)}(c)(x-c)^n}{-(n+1)(x-c)^n} [0 - (x-a)^{n+1}] \\ &= \frac{1}{(n+1)!} f^{(n+1)}(c)(x-a)^{n+1} \end{aligned}$$

Don't forget, when computing $G'(c)$, that G is a function of t with x just a fixed parameter.

Proof 2: We apply Theorem 2.2.10 (the mean value theorem for weighted integrals). If $a < x$, we use the weight function $w(t) = \frac{1}{n!}(x-t)^n$, which is strictly positive for all $a < t < x$. By part (a) this gives

$$\begin{aligned} E_n(a, x) &= \int_a^x \frac{1}{n!} f^{(n+1)}(t) (x-t)^n dt \\ &= f^{(n+1)}(c) \int_a^x \frac{1}{n!} (x-t)^n dt \quad \text{for some } a < c < x \\ &= f^{(n+1)}(c) \left[-\frac{1}{n!} \frac{(x-t)^{n+1}}{n+1} \right]_a^x \\ &= \frac{1}{(n+1)!} f^{(n+1)}(c) (x-a)^{n+1} \end{aligned}$$

If $x < a$, we instead use the weight function $w(t) = \frac{1}{n!}(t-x)^n$, which is strictly positive for all $x < t < a$. This gives

$$\begin{aligned} E_n(a, x) &= \int_a^x \frac{1}{n!} f^{(n+1)}(t) (x-t)^n dt \\ &= -(-1)^n \int_x^a \frac{1}{n!} f^{(n+1)}(t) (t-x)^n dt \\ &= (-1)^{n+1} f^{(n+1)}(c) \int_x^a \frac{1}{n!} (t-x)^n dt \quad \text{for some } x < c < a \\ &= (-1)^{n+1} f^{(n+1)}(c) \left[\frac{1}{n!} \frac{(t-x)^{n+1}}{n+1} \right]_x^a \\ &= \frac{1}{(n+1)!} f^{(n+1)}(c) (-1)^{n+1} (a-x)^{n+1} \\ &= \frac{1}{(n+1)!} f^{(n+1)}(c) (x-a)^{n+1} \end{aligned}$$

Theorem 3.6.11 has provided us with two formulae for the Taylor remainder $E_n(a, x)$. The formula of part (b), $E_n(a, x) = \frac{1}{(n+1)!} f^{(n+1)}(c) (x-a)^{n+1}$, is probably the easiest to use, and the most commonly used, formula for $E_n(a, x)$. The formula of part (a), $E_n(a, x) = \int_a^x \frac{1}{n!} f^{(n+1)}(t) (x-t)^n dt$, while a bit harder to apply, gives a bit better bound than that of part (b) (in the proof of Theorem 3.6.11 we showed that part (b) follows from part (a)). Here is an example in which we use both parts.

✓ Example 3.6.12

In Theorem 3.6.7 we stated that

$$\log(1+x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

for all $-1 < x \leq 1$. But, so far, we have not justified this statement. We do so now, using (both parts of) Theorem 3.6.11. We start by setting $f(x) = \log(1+x)$ and finding the Taylor polynomials $T_n(0, x)$, and the corresponding errors $E_n(0, x)$, for $f(x)$.

$$\begin{array}{ll}
 f(x) = \log(1+x) & f(0) = \log 1 = 0 \\
 f'(x) = \frac{1}{1+x} & f'(0) = 1 \\
 f''(x) = \frac{-1}{(1+x)^2} & f''(0) = -1 \\
 f'''(x) = \frac{2}{(1+x)^3} & f'''(0) = 2 \\
 f^{(4)}(x) = \frac{-2 \times 3}{(1+x)^4} & f^{(4)}(0) = -3! \\
 f^{(5)}(x) = \frac{2 \times 3 \times 4}{(1+x)^5} & f^{(5)}(0) = 4! \\
 \vdots & \vdots \\
 f^{(n)}(x) = \frac{(-1)^{n+1}(n-1)!}{(1+x)^n} & f^{(n)}(0) = (-1)^{n+1}(n-1)!
 \end{array}$$

So the Taylor polynomial of degree n for the function $f(x) = \log(1+x)$, expanded about $a = 0$, is

$$\begin{aligned}
 T_n(0, x) &= f(0) + f'(0)x + \dots + \frac{1}{n!}f^{(n)}(0)x^n \\
 &= x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \frac{1}{5}x^5 + \dots + \frac{(-1)^{n+1}}{n}x^n
 \end{aligned}$$

Theorem 3.6.11 gives us two formulae for the error $E_n(0, x) = f(x) - T_n(0, x)$ made when we approximate $f(x)$ by $T_n(0, x)$. Part (a) of the theorem gives

$$E_n(0, x) = \int_0^x \frac{1}{n!}f^{(n+1)}(t)(x-t)^n dt = (-1)^n \int_0^x \frac{(x-t)^n}{(1+t)^{n+1}} dt$$

and part (b) gives

$$E_n(0, x) = \frac{1}{(n+1)!}f^{(n+1)}(c)x^{n+1} = (-1)^n \frac{1}{n+1} \frac{x^{n+1}}{(1+c)^{n+1}}$$

for some (unknown) c between 0 and x . The statement (S1), that we wish to prove, is equivalent to the statement

$$\lim_{n \rightarrow \infty} E_n(0, x) = 0 \quad \text{for all } -1 < x \leq 1$$

and we will now show that (S2) is true.

The case $x = 0$:

This case is trivial, since, when $x = 0$, $E_n(0, x) = 0$ for all n .

The case $0 < x \leq 1$:

This case is relatively easy to deal with using (Eb). In this case $0 < x \leq 1$, so that the c of (Eb) must be positive and

$$\begin{aligned}
 |E_n(0, x)| &= \frac{1}{n+1} \frac{x^{n+1}}{(1+c)^{n+1}} \\
 &\leq \frac{1}{n+1} \frac{1^{n+1}}{(1+0)^{n+1}} \\
 &= \frac{1}{n+1}
 \end{aligned}$$

converges to zero as $n \rightarrow \infty$.

The case $-1 < x < 0$:

When $-1 < x < 0$ is close to -1 , (Eb) is not sufficient to show that (S2) is true. To see this, let's consider the example $x = -0.8$. All we know about the c of (Eb) is that it has to be between 0 and -0.8 . For example, (Eb) certainly allows c to be -0.6 and then

$$\begin{aligned} & \left| (-1)^n \frac{1}{n+1} \frac{x^{n+1}}{(1+c)^{n+1}} \right|_{\substack{x=-0.8 \\ c=-0.6}} \\ &= \frac{1}{n+1} \frac{0.8^{n+1}}{(1-0.6)^{n+1}} \\ &= \frac{1}{n+1} 2^{n+1} \end{aligned}$$

goes to $+\infty$ as $n \rightarrow \infty$.

Note that, while this does tell us that (Eb) is not sufficient to prove (S2), when x is close to -1 , it does not also tell us that $\lim_{n \rightarrow \infty} |E_n(0, -0.8)| = +\infty$ (which would imply that (S2) is false) — c could equally well be -0.2 and then

$$\begin{aligned} & \left| (-1)^n \frac{1}{n+1} \frac{x^{n+1}}{(1+c)^{n+1}} \right|_{\substack{x=-0.8 \\ c=-0.2}} \\ &= \frac{1}{n+1} \frac{0.8^{n+1}}{(1-0.2)^{n+1}} \\ &= \frac{1}{n+1} \end{aligned}$$

goes to 0 as $n \rightarrow \infty$.

We'll now use (Ea) (which has the advantage of not containing any unknown free parameter c) to verify (S2) when $-1 < x < 0$. Rewrite the right hand side of (Ea)

$$\begin{aligned} (-1)^n \int_0^x \frac{(x-t)^n}{(1+t)^{n+1}} dt &= - \int_x^0 \frac{(t-x)^n}{(1+t)^{n+1}} dt \\ &= - \int_0^{-x} \frac{s^n}{(1+x+s)^{n+1}} ds \quad s = t-x, \quad ds = dt \end{aligned}$$

The exact evaluation of this integral is very messy and not very illuminating. Instead, we bound it. Note that, for $1+x > 0$,

$$\begin{aligned} \frac{d}{ds} \left(\frac{s}{1+x+s} \right) &= \frac{d}{ds} \left(\frac{1+x+s-(1+x)}{1+x+s} \right) \\ &= \frac{d}{ds} \left(1 - \frac{1+x}{1+x+s} \right) \\ &= \frac{1+x}{(1+x+s)^2} > 0 \end{aligned}$$

so that $\frac{s}{1+x+s}$ increases as s increases. Consequently, the biggest value that $\frac{s}{1+x+s}$ takes on the domain of integration $0 \leq s \leq -x = |x|$ is

$$\frac{s}{1+x+s} \Big|_{s=-x} = -x = |x|$$

and the integrand

$$\begin{aligned} 0 \leq \frac{s^n}{[1+x+s]^{n+1}} &= \left(\frac{s}{1+x+s} \right)^n \frac{1}{1+x+s} \\ &\leq \frac{|x|^n}{1+x+s} \end{aligned}$$

Consequently,

$$\begin{aligned}
 |E_n(0, x)| &= \left| (-1)^n \int_0^x \frac{(x-t)^n}{(1+t)^{n+1}} dt \right| \\
 &= \int_0^{-x} \frac{s^n}{[1+x+s]^{n+1}} ds \\
 &\leq |x|^n \int_0^{-x} \frac{1}{1+x+s} ds \\
 &= |x|^n \left[\log(1+x+s) \right]_{s=0}^{s=-x} \\
 &= |x|^n [-\log(1+x)]
 \end{aligned}$$

converges to zero as $n \rightarrow \infty$ for each fixed $-1 < x < 0$.

So we have verified (S2), as desired.

As we said above, Theorem 3.6.11 gave the two most commonly used formulae for the Taylor remainder. Here are some less commonly used, but occasionally useful, formulae.

Theorem 3.6.13 More formulae for the Taylor remainder

1. If $G(t)$ is differentiable¹¹ and $G'(c)$ is nonzero for all c strictly between a and x , then the Taylor remainder

$$E_n(a, x) = \frac{1}{n!} f^{(n+1)}(c) \frac{G(x) - G(a)}{G'(c)} (x - c)^n$$

for some c strictly between a and x .

2. (Cauchy form)

$$E_n(a, x) = \frac{1}{n!} f^{(n+1)}(c) (x - c)^n (x - a)$$

for some c strictly between a and x .

Proof

As in the proof of Theorem 3.6.11, we define

$$S(t) = f(x) - f(t) - f'(t)(x-t) - \frac{1}{2}f''(t)(x-t)^2 - \dots - \frac{1}{n!}f^{(n)}(t)(x-t)^n$$

and observe that $E_n(a, x) = S(a)$ and $S(x) = 0$ and $S'(t) = -\frac{1}{n!}f^{(n+1)}(t)(x-t)^n$.

1. Recall that the generalised mean-value theorem, which is Theorem 3.4.38 in the CLP-1 text, says that

$$\frac{F(b) - F(a)}{G(b) - G(a)} = \frac{F'(c)}{G'(c)}$$

for some c strictly between a and b . We apply this theorem with $b = x$ and $F(t) = S(t)$. This gives

$$\begin{aligned}
 E_n(a, x) &= -[S(x) - S(a)] = -\frac{S'(c)}{G'(c)} [G(x) - G(a)] \\
 &= -\frac{-\frac{1}{n!}f^{(n+1)}(c)(x-c)^n}{G'(c)} [G(x) - G(a)] \\
 &= \frac{1}{n!}f^{(n+1)}(c) \frac{G(x) - G(a)}{G'(c)} (x - c)^n
 \end{aligned}$$

2. Apply part (a) with $G(x) = x$. This gives

$$\begin{aligned} E_n(a, x) &= \frac{1}{n!} f^{(n+1)}(c) \frac{x-a}{1} (x-c)^n \\ &= \frac{1}{n!} f^{(n+1)}(c) (x-c)^n (x-a) \end{aligned}$$

for some c strictly between a and b .

✓ Example 3.6.14 Example 3.6.12, continued

In Example 3.6.12 we verified that

$$\log(1+x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

for all $-1 < x \leq 1$. There we used the Lagrange form,

$$E_n(a, x) = \frac{1}{(n+1)!} f^{(n+1)}(c) (x-a)^{n+1}$$

for the Taylor remainder to verify (S1) when $0 \leq x \leq 1$, but we also saw that it is not possible to use the Lagrange form to verify (S1) when x is close to -1 . We instead used the integral form

$$E_n(a, x) = \int_a^x \frac{1}{n!} f^{(n+1)}(t) (x-t)^n dt$$

We will now use the Cauchy form (part (b) of Theorem 3.6.13)

$$E_n(a, x) = \frac{1}{n!} f^{(n+1)}(c) (x-c)^n (x-a)$$

to verify

$$\lim_{n \rightarrow \infty} E_n(0, x) = 0$$

when $-1 < x < 0$. We have already noted that (S2) is equivalent to (S1).

Write $f(x) = \log(1+x)$. We saw in Example 3.6.12 that

$$f^{(n+1)}(x) = \frac{(-1)^n n!}{(1+x)^{n+1}}$$

So, in this example, the Cauchy form is

$$E_n(0, x) = (-1)^n \frac{(x-c)^n x}{(1+c)^{n+1}}$$

for some $x < c < 0$. When $-1 < x < c < 0$,

- c and x are negative and $1+x$, $1+c$ and $c-x$ are (strictly) positive so that

$$\begin{aligned} c(1+x) < 0 &\implies c < -cx \implies c-x < -x-xc = |x|(1+c) \\ &\implies \left| \frac{x-c}{1+c} \right| = \frac{c-x}{1+c} < |x| \end{aligned}$$

so that $\left| \frac{x-c}{1+c} \right|^n < |x|^n$ and

- the distance from -1 to c , namely $c - (-1) = 1+c$ is greater than the distance from -1 to x , namely $x - (-1) = 1+x$, so that $\frac{1}{1+c} < \frac{1}{1+x}$.

So, for $-1 < x < c < 0$,

$$|E_n(0, x)| = \left| \frac{x-c}{1+c} \right|^n \frac{|x|}{1+c} < \frac{|x|^{n+1}}{1+c} < \frac{|x|^{n+1}}{1+x}$$

goes to zero as $n \rightarrow \infty$.

Computing with Taylor Series

Taylor series have a great many applications. (Hence their place in this course.) One of the most immediate of these is that they give us an alternate way of computing many functions. For example, the first definition we see for the sine and cosine functions is in terms of triangles. Those definitions, however, do not lend themselves to computing sine and cosine except at very special angles. Armed with power series representations, however, we can compute them to very high precision at any angle. To illustrate this, consider the computation of π — a problem that dates back to the Babylonians.

✓ Example 3.6.15 Computing the number π

There are numerous methods for computing π to any desired degree of accuracy¹². Many of them use the Maclaurin expansion

$$\arctan x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$$

of Theorem 3.6.7. Since $\arctan(1) = \frac{\pi}{4}$, the series gives us a very pretty formula for π :

$$\begin{aligned} \frac{\pi}{4} &= \arctan 1 = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \\ \pi &= 4 \left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots \right) \end{aligned}$$

Unfortunately, this series is not very useful for computing π because it converges so slowly. If we approximate the series by its N^{th} partial sum, then the alternating series test (Theorem 3.3.14) tells us that the error is bounded by the first term we drop. To guarantee that we have 2 decimal digits of π correct, we need to sum about the first 200 terms!

A much better way to compute π using this series is to take advantage of the fact that $\tan \frac{\pi}{6} = \frac{1}{\sqrt{3}}$:

$$\begin{aligned} \pi &= 6 \arctan \left(\frac{1}{\sqrt{3}} \right) = 6 \sum_{n=0}^{\infty} (-1)^n \frac{1}{2n+1} \left(\frac{1}{\sqrt{3}} \right)^{2n+1} \\ &= 2\sqrt{3} \sum_{n=0}^{\infty} (-1)^n \frac{1}{2n+1} \left(\frac{1}{3} \right)^n = 2\sqrt{3} \left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} - \frac{1}{15} + \cdots \right) \end{aligned}$$

Again, this is an alternating series and so (via Theorem 3.3.14) the error we introduce by truncating it is bounded by the first term dropped. For example, if we keep ten terms, stopping at $n = 9$, we get $\pi = 3.141591$ (to 6 decimal places) with an error between zero and

$$\frac{2\sqrt{3}}{21 \times 3^{10}} < 3 \times 10^{-6}$$

In 1699, the English astronomer/mathematician Abraham Sharp (1653–1742) used 150 terms of this series to compute 72 digits of π — by hand!

This is just one of very many ways to compute π . Another one, which still uses the Maclaurin expansion of $\arctan x$, but is much more efficient, is

$$\pi = 16 \arctan \frac{1}{5} - 4 \arctan \frac{1}{239}$$

This formula was used by John Machin in 1706 to compute π to 100 decimal digits — again, by hand.

Power series also give us access to new functions which might not be easily expressed in terms of the functions we have been introduced to so far. The following is a good example of this.

✓ Example 3.6.16 Error function

The *error function*

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

is used in computing “bell curve” probabilities. The indefinite integral of the integrand e^{-t^2} cannot be expressed in terms of standard functions. But we can still evaluate the integral to within any desired degree of accuracy by using the Taylor expansion of the exponential. Start with the Maclaurin series for e^x :

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$$

and then substitute $x = -t^2$ into this:

$$e^{-t^2} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} t^{2n}$$

We can then apply Theorem 3.5.13 to integrate term-by-term:

$$\begin{aligned} \operatorname{erf}(x) &= \frac{2}{\sqrt{\pi}} \int_0^x \left(\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)n!} \right) dx \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+2}}{(2n+2)n!} \end{aligned}$$

For example, for the bell curve, the probability of being within one standard deviation of the mean¹³, is

$$\begin{aligned} \operatorname{erf}\left(\frac{1}{\sqrt{2}}\right) &= \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n \left(\frac{1}{\sqrt{2}}\right)^{2n+1}}{(2n+1)n!} = \frac{2}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)2^n n!} \\ &= \sqrt{\frac{2}{\pi}} \left(1 - \frac{1}{3 \times 2} + \frac{1}{5 \times 2^2 \times 2} - \frac{1}{7 \times 2^3 \times 3!} + \frac{1}{9 \times 2^4 \times 4!} - \dots \right) \end{aligned}$$

This is yet another alternating series. If we keep five terms, stopping at $n = 4$, we get 0.68271 (to 5 decimal places) with, by Theorem 3.3.14 again, an error between zero and the first dropped term, which is minus

$$\sqrt{\frac{2}{\pi}} \frac{1}{11 \times 2^5 \times 5!} < 2 \times 10^{-5}$$

✓ Example 3.6.17 Two nice series

Evaluate

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n3^n} \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{n3^n}$$

Solution. There are not very many series that can be easily evaluated exactly. But occasionally one encounters a series that can be evaluated simply by realizing that it is exactly one of the series in Theorem 3.6.7, just with a specific value of x . The left hand given series is

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \frac{1}{3^n} = \frac{1}{3} - \frac{1}{2} \frac{1}{3^2} + \frac{1}{3} \frac{1}{3^3} - \frac{1}{4} \frac{1}{3^4} + \dots$$

The series in Theorem 3.6.7 that this most closely resembles is

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

Indeed

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \frac{1}{3^n} &= \frac{1}{3} - \frac{1}{2} \frac{1}{3^2} + \frac{1}{3} \frac{1}{3^3} - \frac{1}{4} \frac{1}{3^4} + \dots \\ &= \left[x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} - \dots \right]_{x=\frac{1}{3}} \\ &= \left[\log(1+x) \right]_{x=\frac{1}{3}} \\ &= \log \frac{4}{3} \end{aligned}$$

The right hand series above differs from the left hand series above only that the signs of the left hand series alternate while those of the right hand series do not. We can flip every second sign in a power series just by using a negative x .

$$\begin{aligned} \left[\log(1+x) \right]_{x=-\frac{1}{3}} &= \left[x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} - \dots \right]_{x=-\frac{1}{3}} \\ &= -\frac{1}{3} - \frac{1}{2} \frac{1}{3^2} - \frac{1}{3} \frac{1}{3^3} - \frac{1}{4} \frac{1}{3^4} + \dots \end{aligned}$$

which is exactly minus the desired right hand series. So

$$\sum_{n=1}^{\infty} \frac{1}{n3^n} = -\left[\log(1+x) \right]_{x=-\frac{1}{3}} = -\log \frac{2}{3} = \log \frac{3}{2}$$

✓ Example 3.6.18 Finding a derivative from a series

Let $f(x) = \sin(2x^3)$. Find $f^{(15)}(0)$, the fifteenth derivative of f at $x = 0$.

Solution: This is a bit of a trick question. We could of course use the product and chain rules to directly apply fifteen derivatives and then set $x = 0$, but that would be extremely tedious¹⁴. There is a much more efficient approach that exploits two pieces of knowledge that we have.

- From equation 3.6.2, we see that the coefficient of $(x-a)^n$ in the Taylor series of $f(x)$ with expansion point a is exactly $\frac{1}{n!} f^{(n)}(a)$. So $f^{(n)}(a)$ is exactly $n!$ times the coefficient of $(x-a)^n$ in the Taylor series of $f(x)$ with expansion point a .
- We know, or at least can easily find, the Taylor series for $\sin(2x^3)$.

Let's apply that strategy.

- First, we know that, for all y ,

$$\sin y = y - \frac{1}{3!}y^3 + \frac{1}{5!}y^5 - \dots$$

- Just substituting $y = 2x^3$, we have

$$\begin{aligned} \sin(2x^3) &= 2x^3 - \frac{1}{3!}(2x^3)^3 + \frac{1}{5!}(2x^3)^5 - \dots \\ &= 2x^3 - \frac{8}{3!}x^9 + \frac{2^5}{5!}x^{15} - \dots \end{aligned}$$

- So the coefficient of x^{15} in the Taylor series of $f(x) = \sin(2x^3)$ with expansion point $a = 0$ is $\frac{2^5}{5!}$

and we have

$$f^{(15)}(0) = 15! \times \frac{2^5}{5!} = 348,713,164,800$$

Optional — Linking e^x with trigonometric functions

Let us return to the observation that we made earlier about the Maclaurin series for sine, cosine and the exponential functions:

$$\begin{aligned}\cos x + \sin x &= 1 + x - \frac{1}{2!}x^2 - \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \frac{1}{5!}x^5 - \dots \\ e^x &= 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \frac{1}{5!}x^5 + \dots\end{aligned}$$

We see that these series are identical except for the differences in the signs of the coefficients. Let us try to make them look even more alike by introducing extra constants A , B and q into the equations. Consider

$$\begin{aligned}A \cos x + B \sin x &= A + Bx - \frac{A}{2!}x^2 - \frac{B}{3!}x^3 + \frac{A}{4!}x^4 + \frac{B}{5!}x^5 - \dots \\ e^{qx} &= 1 + qx + \frac{q^2}{2!}x^2 + \frac{q^3}{3!}x^3 + \frac{q^4}{4!}x^4 + \frac{q^5}{5!}x^5 + \dots\end{aligned}$$

Let's try to choose A , B and q so that these two expressions are equal. To do so we must make sure that the coefficients of the various powers of x agree. Looking just at the coefficients of x^0 and x^1 , we see that we need

$$A = 1 \quad \text{and} \quad B = q$$

Substituting this into our expansions gives

$$\begin{aligned}\cos x + q \sin x &= 1 + qx - \frac{1}{2!}x^2 - \frac{q}{3!}x^3 + \frac{1}{4!}x^4 + \frac{q}{5!}x^5 - \dots \\ e^{qx} &= 1 + qx + \frac{q^2}{2!}x^2 + \frac{q^3}{3!}x^3 + \frac{q^4}{4!}x^4 + \frac{q^5}{5!}x^5 + \dots\end{aligned}$$

Now the coefficients of x^0 and x^1 agree, but the coefficient of x^2 tells us that we need q to be a number so that $q^2 = -1$, or

$$q = \sqrt{-1}$$

We know that no such *real* number q exists. But for the moment let us see what happens if we just assume¹⁷ that we can find q so that $q^2 = -1$. Then we will have that

$$q^3 = -q \quad q^4 = 1 \quad q^5 = q \quad \dots$$

so that the series for $\cos x + q \sin x$ and e^{qx} are identical. That is

$$e^{qx} = \cos x + q \sin x$$

If we now write this with the more usual notation $q = \sqrt{-1} = i$ we arrive at what is now known as Euler's formula

Equation 3.6.20

$$e^{ix} = \cos x + i \sin x$$

Euler's proof of this formula (in 1740) was based on Maclaurin expansions (much like our explanation above). Euler's formula¹⁸ is widely regarded as one of the most important and beautiful in all of mathematics.

Of course having established Euler's formula one can find slicker demonstrations. For example, let

$$f(x) = e^{-ix} (\cos x + i \sin x)$$

Differentiating (with product and chain rules and the fact that $i^2 = -1$) gives us

$$\begin{aligned}f'(x) &= -ie^{-ix} (\cos x + i \sin x) + e^{-ix} (-\sin x + i \cos x) \\ &= 0\end{aligned}$$

Since the derivative is zero, the function $f(x)$ must be a constant. Setting $x = 0$ tells us that

$$f(0) = e^0 (\cos 0 + i \sin 0) = 1.$$

Hence $f(x) = 1$ for all x . Rearranging then arrives at

$$e^{ix} = \cos x + i \sin x$$

as required.

Substituting $x = \pi$ into Euler's formula we get Euler's identity

$$e^{i\pi} = -1$$

which is more often stated

Equation 3.6.21 Euler's identity

$$e^{i\pi} + 1 = 0$$

which links the 5 most important constants in mathematics, 1, 0, π , e and $\sqrt{-1}$.

Evaluating Limits using Taylor Expansions

Taylor polynomials provide a good way to understand the behaviour of a function near a specified point and so are useful for evaluating complicated limits. Here are some examples.

✓ Example 3.6.22 A simple limit from a Taylor expansion

In this example, we'll start with a relatively simple limit, namely

$$\lim_{x \rightarrow 0} \frac{\sin x}{x}$$

The first thing to notice about this limit is that, as x tends to zero, both the numerator, $\sin x$, and the denominator, x , tend to 0. So we may not evaluate the limit of the ratio by simply dividing the limits of the numerator and denominator. To find the limit, or show that it does not exist, we are going to have to exhibit a cancellation between the numerator and the denominator. Let's start by taking a closer look at the numerator. By Example 3.6.6,

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots$$

Consequently¹⁹

$$\frac{\sin x}{x} = 1 - \frac{1}{3!}x^2 + \frac{1}{5!}x^4 - \dots$$

Every term in this series, except for the very first term, is proportional to a strictly positive power of x . Consequently, as x tends to zero, all terms in this series, except for the very first term, tend to zero. In fact the sum of all terms, starting with the second term, also tends to zero. That is,

$$\lim_{x \rightarrow 0} \left[-\frac{1}{3!}x^2 + \frac{1}{5!}x^4 - \dots \right] = 0$$

We won't justify that statement here, but it will be justified in the following (optional) subsection. So

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{\sin x}{x} &= \lim_{x \rightarrow 0} \left[1 - \frac{1}{3!}x^2 + \frac{1}{5!}x^4 - \dots \right] \\ &= 1 + \lim_{x \rightarrow 0} \left[-\frac{1}{3!}x^2 + \frac{1}{5!}x^4 - \dots \right] \\ &= 1 \end{aligned}$$

The limit in the previous example can also be evaluated relatively easily using l'Hôpital's rule²⁰. While the following limit can also, in principal, be evaluated using l'Hôpital's rule, it is much more efficient to use Taylor series²¹.

✓ Example 3.6.23 A not so easy limit made easier

In this example we evaluate

$$\lim_{x \rightarrow 0} \frac{\arctan x - x}{\sin x - x}$$

Once again, the first thing to notice about this limit is that, as x tends to zero, the numerator tends to $\arctan 0 - 0$, which is 0, and the denominator tends to $\sin 0 - 0$, which is also 0. So we may not evaluate the limit of the ratio by simply dividing the limits of the numerator and denominator. Again, to find the limit, or show that it does not exist, we are going to have to exhibit a cancellation between the numerator and the denominator. To get a more detailed understanding of the behaviour of the numerator and denominator near $x = 0$, we find their Taylor expansions. By Example 3.5.21,

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \dots$$

so the numerator

$$\arctan x - x = -\frac{x^3}{3} + \frac{x^5}{5} - \dots$$

By Example 3.6.6,

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots$$

so the denominator

$$\sin x - x = -\frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots$$

and the ratio

$$\frac{\arctan x - x}{\sin x - x} = \frac{-\frac{x^3}{3} + \frac{x^5}{5} - \dots}{-\frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots}$$

Notice that every term in both the numerator and the denominator contains a common factor of x^3 , which we can cancel out.

$$\frac{\arctan x - x}{\sin x - x} = \frac{-\frac{1}{3} + \frac{x^2}{5} - \dots}{-\frac{1}{3!} + \frac{1}{5!}x^2 - \dots}$$

As x tends to zero,

- the numerator tends to $-\frac{1}{3}$, which is not 0, and
- the denominator tends to $-\frac{1}{3!} = -\frac{1}{6}$, which is also not 0.

so we may now legitimately evaluate the limit of the ratio by simply dividing the limits of the numerator and denominator.

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{\arctan x - x}{\sin x - x} &= \lim_{x \rightarrow 0} \frac{-\frac{1}{3} + \frac{x^2}{5} - \dots}{-\frac{1}{3!} + \frac{1}{5!}x^2 - \dots} \\ &= \frac{\lim_{x \rightarrow 0} \left[-\frac{1}{3} + \frac{x^2}{5} - \dots \right]}{\lim_{x \rightarrow 0} \left[-\frac{1}{3!} + \frac{1}{5!}x^2 - \dots \right]} \\ &= \frac{-\frac{1}{3}}{-\frac{1}{3!}} \\ &= 2 \end{aligned}$$

Optional — The Big O Notation

In Example 3.6.22 we used, without justification²², that, as x tends to zero, not only does every term in

$$\frac{\sin x}{x} - 1 = -\frac{1}{3!}x^2 + \frac{1}{5!}x^4 - \dots = \sum_{n=1}^{\infty} (-1)^n \frac{1}{(2n+1)!} x^{2n}$$

converge to zero, but in fact the sum of all infinitely many terms also converges to zero. We did something similar twice in Example 3.6.23; once in computing the limit of the numerator and once in computing the limit of the denominator.

We'll now develop some machinery that provides the justification. We start by recalling, from equation 3.6.1, that if, for some natural number n , the function $f(x)$ has $n+1$ derivatives near the point a , then

$$f(x) = T_n(x) + E_n(x)$$

where

$$T_n(x) = f(a) + f'(a)(x-a) + \dots + \frac{1}{n!}f^{(n)}(a)(x-a)^n$$

is the Taylor polynomial of degree n for the function $f(x)$ and expansion point a and

$$E_n(x) = f(x) - T_n(x) = \frac{1}{(n+1)!}f^{(n+1)}(c)(x-a)^{n+1}$$

is the error introduced when we approximate $f(x)$ by the polynomial $T_n(x)$. Here c is some unknown number between a and x . As c is not known, we do not know exactly what the error $E_n(x)$ is. But that is usually not a problem.

In the present context²³ we are interested in taking the limit as $x \rightarrow a$. So we are only interested in x -values that are very close to a , and because c lies between x and a , c is also very close to a . Now, as long as $f^{(n+1)}(x)$ is continuous at a , as $x \rightarrow a$, $f^{(n+1)}(c)$ must approach $f^{(n+1)}(a)$ which is some finite value. This, in turn, means that there must be constants $M, D > 0$ such that $|f^{(n+1)}(c)| \leq M$ for all c 's within a distance D of a . If so, there is another constant C (namely $\frac{M}{(n+1)!}$) such that

$$|E_n(x)| \leq C|x-a|^{n+1} \quad \text{whenever } |x-a| \leq D$$

There is some notation for this behaviour.

Definition 3.6.24 Big O

Let a and m be real numbers. We say that the function “ $g(x)$ is of order $|x-a|^m$ near a ” and we write $g(x) = O(|x-a|^m)$ if there exist constants²⁴ $C, D > 0$ such that

$$|g(x)| \leq C|x-a|^m \quad \text{whenever } |x-a| \leq D \quad (\star)$$

Whenever $O(|x-a|^m)$ appears in an algebraic expression, it just stands for some (unknown) function $g(x)$ that obeys (\star) . This is called “big O” notation.

How should we parse the big O notation when we see it? Consider the following

$$g(x) = O(|x-3|^2)$$

First of all, we know from the definition that the notation only tells us something about $g(x)$ for x near the point a . The equation above contains “ $O(|x-3|^2)$ ” which tells us something about what the function looks like when x is close to 3. Further, because it is “ $|x-3|$ ” squared, it says that the graph of the function lies below a parabola $y = C(x-3)^2$ and above a parabola $y = -C(x-3)^2$ near $x = 3$. The notation doesn't tell us anything more than this — we don't know, for example, that the graph of $g(x)$ is concave up or concave down. It also tells us that Taylor expansion of $g(x)$ around $x = 3$ does not contain any constant or linear term — the first nonzero term in the expansion is of degree at least two. For example, all of the following functions are $O(|x-3|^2)$.

$$5(x-3)^2 + 6(x-3)^3, \quad -7(x-3)^2 - 8(x-3)^4, \quad (x-3)^3, \quad (x-3)^{\frac{5}{2}}$$

In the next few examples we will rewrite a few of the Taylor polynomials that we know using this big O notation.

✓ Example 3.6.25 Sine and the big O

Let $f(x) = \sin x$ and $a = 0$. Then

$$\begin{array}{llll} f(x) = \sin x & f'(x) = \cos x & f''(x) = -\sin x & f^{(3)}(x) = -\cos x \\ f(0) = 0 & f'(0) = 1 & f''(0) = 0 & f^{(3)}(0) = -1 \\ f^{(4)}(x) = \sin x & \dots & & \\ f^{(4)}(0) = 0 & \dots & & \end{array}$$

and the pattern repeats. So every derivative is plus or minus either sine or cosine and, as we saw in previous examples, this makes analysing the error term for the sine and cosine series quite straightforward. In particular, $|f^{(n+1)}(c)| \leq 1$ for all real numbers c and all natural numbers n . So the Taylor polynomial of, for example, degree 3 and its error term are

$$\begin{aligned} \sin x &= x - \frac{1}{3!}x^3 + \frac{\cos c}{5!}x^5 \\ &= x - \frac{1}{3!}x^3 + O(|x|^5) \end{aligned}$$

under Definition 3.6.24, with $C = \frac{1}{5!}$ and any $D > 0$. Similarly, for any natural number n ,

Equation 3.6.26

$$\begin{aligned} \sin x &= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots + (-1)^n \frac{1}{(2n+1)!}x^{2n+1} + O(|x|^{2n+3}) \\ \cos x &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \dots + (-1)^n \frac{1}{(2n)!}x^{2n} + O(|x|^{2n+2}) \end{aligned}$$

When we studied the error in the expansion of the exponential function (way back in optional Example 3.6.8), we had to go to some length to understand the behaviour of the error term well enough to prove convergence for all numbers x . However, in the big O notation, we are free to assume that x is close to 0. Furthermore we do not need to derive an explicit bound on the size of the coefficient C . This makes it quite a bit easier to verify that the big O notation is correct.

✓ Example 3.6.27 Exponential and the big O

Let n be any natural number. Since $\frac{d}{dx}e^x = e^x$, we know that $\frac{d^k}{dx^k}\{e^x\} = e^x$ for every integer $k \geq 0$. Thus

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + \frac{e^c}{(n+1)!}x^{n+1}$$

for some c between 0 and x . If, for example, $|x| \leq 1$, then $|e^c| \leq e$, so that the error term

$$\left| \frac{e^c}{(n+1)!}x^{n+1} \right| \leq C|x|^{n+1} \quad \text{with } C = \frac{e}{(n+1)!} \quad \text{whenever } |x| \leq 1$$

So, under Definition 3.6.24, with $C = \frac{e}{(n+1)!}$ and $D = 1$,

Equation 3.6.28

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + O(|x|^{n+1})$$

You can see that, because we only have to consider x 's that are close to the expansion point (in this example, 0) it is relatively easy to derive the bounds that are required to justify the use of the big O notation.

✓ Example 3.6.29 Logarithms and the big O

Let $f(x) = \log(1+x)$ and $a = 0$. Then

$$\begin{aligned}
 f'(x) &= \frac{1}{1+x} & f''(x) &= -\frac{1}{(1+x)^2} & f^{(3)}(x) &= \frac{2}{(1+x)^3} \\
 f'(0) &= 1 & f''(0) &= -1 & f^{(3)}(0) &= 2 \\
 f^{(4)}(x) &= -\frac{2 \times 3}{(1+x)^4} & f^{(5)}(x) &= \frac{2 \times 3 \times 4}{(1+x)^5} \\
 f^{(4)}(0) &= -3! & f^{(5)}(0) &= 4!
 \end{aligned}$$

We can see a pattern for $f^{(n)}(x)$ forming here — $f^{(n)}(x)$ is a sign times a ratio with

- the sign being + when n is odd and being - when n is even. So the sign is $(-1)^{n-1}$.
- The denominator is $(1+x)^n$.
- The numerator²⁵ is the product $2 \times 3 \times 4 \times \dots \times (n-1) = (n-1)!$.

Thus²⁶, for any natural number n ,

$$\begin{aligned}
 f^{(n)}(x) &= (-1)^{n-1} \frac{(n-1)!}{(1+x)^n} && \text{which means that} \\
 \frac{1}{n!} f^{(n)}(0) x^n &= (-1)^{n-1} \frac{(n-1)!}{n!} x^n = (-1)^{n-1} \frac{x^n}{n}
 \end{aligned}$$

so

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots + (-1)^{n-1} \frac{x^n}{n} + E_n(x)$$

with

$$E_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(c) (x-a)^{n+1} = \frac{1}{n+1} \cdot \frac{(-1)^n}{(1+c)^{n+1}} \cdot x^{n+1}$$

If we choose, for example $D = \frac{1}{2}$, then²⁷ for any x obeying $|x| \leq D = \frac{1}{2}$, we have $|c| \leq \frac{1}{2}$ and $|1+c| \geq \frac{1}{2}$ so that

$$|E_n(x)| \leq \frac{1}{(n+1)(1/2)^{n+1}} |x|^{n+1} = O(|x|^{n+1})$$

under Definition 3.6.24, with $C = \frac{2^{n+1}}{n+1}$ and $D = \frac{1}{2}$. Thus we may write

Equation 3.6.30

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots + (-1)^{n-1} \frac{x^n}{n} + O(|x|^{n+1})$$

Remark 3.6.31

The big O notation has a few properties that are useful in computations and taking limits. All follow immediately from Definition 3.6.24.

1. If $p > 0$, then

$$\lim_{x \rightarrow 0} O(|x|^p) = 0$$

2. For any real numbers p and q ,

$$O(|x|^p) O(|x|^q) = O(|x|^{p+q})$$

(This is just because $C|x|^p \times C'|x|^q = (CC')|x|^{p+q}$.) In particular,

$$ax^m O(|x|^p) = O(|x|^{p+m})$$

for any constant a and any integer m .

3. For any real numbers p and q ,

$$O(|x|^p) + O(|x|^q) = O(|x|^{\min\{p,q\}})$$

(For example, if $p = 2$ and $q = 5$, then $C|x|^2 + C'|x|^5 = (C + C'|x|^3)|x|^2 \leq (C + C')|x|^2$ whenever $|x| \leq 1$.)

4. For any real numbers p and q with $p > q$, any function which is $O(|x|^p)$ is also $O(|x|^q)$ because $C|x|^p = C|x|^{p-q}|x|^q \leq C|x|^q$ whenever $|x| \leq 1$.
5. All of the above observations also hold for more general expressions with $|x|$ replaced by $|x - a|$, i.e. for $O(|x - a|^p)$. The only difference being in (a) where we must take the limit as $x \rightarrow a$ instead of $x \rightarrow 0$.

Optional — Evaluating Limits Using Taylor Expansions — More Examples

✓ Example 3.6.32 Example 3.6.22 revisited

In this example, we'll return to the limit

$$\lim_{x \rightarrow 0} \frac{\sin x}{x}$$

of Example 3.6.22 and treat it more carefully. By Example 3.6.25,

$$\sin x = x - \frac{1}{3!}x^3 + O(|x|^5)$$

That is, for small x , $\sin x$ is the same as $x - \frac{1}{3!}x^3$, up to an error that is bounded by some constant times $|x|^5$. So, dividing by x , $\frac{\sin x}{x}$ is the same as $1 - \frac{1}{3!}x^2$, up to an error that is bounded by some constant times x^4 — see Remark 3.6.31(b). That is

$$\frac{\sin x}{x} = 1 - \frac{1}{3!}x^2 + O(x^4)$$

But any function that is bounded by some constant times x^4 (for all x smaller than some constant $D > 0$) necessarily tends to 0 as $x \rightarrow 0$ — see Remark 3.6.31(a). Thus

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = \lim_{x \rightarrow 0} \left[1 - \frac{1}{3!}x^2 + O(x^4) \right] = \lim_{x \rightarrow 0} \left[1 - \frac{1}{3!}x^2 \right] = 1$$

Reviewing the above computation, we see that we did a little more work than we had to. It wasn't necessary to keep track of the $-\frac{1}{3!}x^3$ contribution to $\sin x$ so carefully. We could have just said that

$$\sin x = x + O(|x|^3)$$

so that

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = \lim_{x \rightarrow 0} \frac{x + O(|x|^3)}{x} = \lim_{x \rightarrow 0} [1 + O(x^2)] = 1$$

We'll spend a little time in the later, more complicated, examples learning how to choose the number of terms we keep in our Taylor expansions so as to make our computations as efficient as possible.

✓ Example 3.6.33 Practicing using Taylor polynomials for limits

In this example, we'll use the Taylor polynomial of Example 3.6.29 to evaluate $\lim_{x \rightarrow 0} \frac{\log(1+x)}{x}$ and $\lim_{x \rightarrow 0} (1+x)^{a/x}$. The Taylor expansion of equation 3.6.30 with $n = 1$ tells us that

$$\log(1+x) = x + O(|x|^2)$$

That is, for small x , $\log(1+x)$ is the same as x , up to an error that is bounded by some constant times x^2 . So, dividing by x , $\frac{1}{x}\log(1+x)$ is the same as 1, up to an error that is bounded by some constant times $|x|$. That is

$$\frac{1}{x}\log(1+x) = 1 + O(|x|)$$

But any function that is bounded by some constant times $|x|$, for all x smaller than some constant $D > 0$, necessarily tends to 0 as $x \rightarrow 0$. Thus

$$\lim_{x \rightarrow 0} \frac{\log(1+x)}{x} = \lim_{x \rightarrow 0} \frac{x + O(|x|^2)}{x} = \lim_{x \rightarrow 0} [1 + O(|x|)] = 1$$

We can now use this limit to evaluate

$$\lim_{x \rightarrow 0} (1+x)^{a/x}.$$

Now, we could either evaluate the limit of the logarithm of this expression, or we can carefully rewrite the expression as $e^{(\text{something})}$. Let us do the latter.

$$\begin{aligned} \lim_{x \rightarrow 0} (1+x)^{a/x} &= \lim_{x \rightarrow 0} e^{\frac{a}{x} \log(1+x)} \\ &= \lim_{x \rightarrow 0} e^{\frac{a}{x} [x + O(|x|^2)]} \\ &= \lim_{x \rightarrow 0} e^{a + O(|x|)} = e^a \end{aligned}$$

Here we have used that if $F(x) = O(|x|^2)$ then $\frac{a}{x}F(x) = O(x)$ — see Remark 3.6.31(b). We have also used that the exponential is continuous — as x tends to zero, the exponent of $e^{a+O(|x|)}$ tends to a so that $e^{a+O(|x|)}$ tends to e^a — see Remark 3.6.31(a).

✓ Example 3.6.34 A difficult limit

In this example, we'll evaluate²⁸ the harder limit

$$\lim_{x \rightarrow 0} \frac{\cos x - 1 + \frac{1}{2}x \sin x}{[\log(1+x)]^4}$$

The first thing to notice about this limit is that, as x tends to zero, the numerator

$$\cos x - 1 + \frac{1}{2}x \sin x \rightarrow \cos 0 - 1 + \frac{1}{2} \cdot 0 \cdot \sin 0 = 0$$

and the denominator

$$[\log(1+x)]^4 \rightarrow [\log(1+0)]^4 = 0$$

too. So both the numerator and denominator tend to zero and we may not simply evaluate the limit of the ratio by taking the limits of the numerator and denominator and dividing.

To find the limit, or show that it does not exist, we are going to have to exhibit a cancellation between the numerator and the denominator. To develop a strategy for evaluating this limit, let's do a “little scratch work”, starting by taking a closer look at the denominator. By Example 3.6.29,

$$\log(1+x) = x + O(x^2)$$

This tells us that $\log(1+x)$ looks a lot like x for very small x . So the denominator $[x + O(x^2)]^4$ looks a lot like x^4 for very small x . Now, what about the numerator?

- If the numerator looks like some constant times x^p with $p > 4$, for very small x , then the ratio will look like the constant times $\frac{x^p}{x^4} = x^{p-4}$ and, as $p - 4 > 0$, will tend to 0 as x tends to zero.
- If the numerator looks like some constant times x^p with $p < 4$, for very small x , then the ratio will look like the constant times $\frac{x^p}{x^4} = x^{p-4}$ and will, as $p - 4 < 0$, tend to infinity, and in particular diverge, as x tends to zero.
- If the numerator looks like Cx^4 , for very small x , then the ratio will look like $\frac{Cx^4}{x^4} = C$ and will tend to C as x tends to zero.

The moral of the above “scratch work” is that we need to know the behaviour of the numerator, for small x , up to order x^4 . Any contributions of order x^p with $p > 4$ may be put into error terms $O(|x|^p)$.

Now we are ready to evaluate the limit. Because the expressions are a little involved, we will simplify the numerator and denominator separately and then put things together. Using the expansions we developed in Example 3.6.25, the numerator,

$$\begin{aligned} \cos x - 1 + \frac{1}{2}x \sin x &= \left(1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + O(|x|^6)\right) \\ &\quad - 1 + \frac{x}{2} \left(x - \frac{1}{3!}x^3 + O(|x|^5)\right) \quad \text{expand} \\ &= \left(\frac{1}{24} - \frac{1}{12}\right)x^4 + O(|x|^6) + \frac{x}{2}O(|x|^5) \end{aligned}$$

Then by Remark 3.6.31(b)

$$= -\frac{1}{24}x^4 + O(|x|^6) + O(|x|^6)$$

and now by Remark 3.6.31(c)

$$= -\frac{1}{24}x^4 + O(|x|^6)$$

Similarly, using the expansion that we developed in Example 3.6.29,

$$\begin{aligned} [\log(1+x)]^4 &= [x + O(|x|^2)]^4 \\ &= [x + xO(|x|)]^4 \quad \text{by Remark 3.6.31(b)} \\ &= x^4[1 + O(|x|)]^4 \end{aligned}$$

Now put these together and take the limit as $x \rightarrow 0$:

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{\cos x - 1 + \frac{1}{2}x \sin x}{[\log(1+x)]^4} &= \lim_{x \rightarrow 0} \frac{-\frac{1}{24}x^4 + O(|x|^6)}{x^4[1 + O(|x|)]^4} \\ &= \lim_{x \rightarrow 0} \frac{-\frac{1}{24}x^4 + x^4O(|x|^2)}{x^4[1 + O(|x|)]^4} \quad \text{by Remark 3.6.31(b)} \\ &= \lim_{x \rightarrow 0} \frac{-\frac{1}{24} + O(|x|^2)}{[1 + O(|x|)]^4} \\ &= -\frac{1}{24} \quad \text{by Remark 3.6.31(a)}. \end{aligned}$$

The next two limits have much the same flavour as those above — expand the numerator and denominator to high enough order, do some cancellations and then take the limit. We have increased the difficulty a little by introducing “expansions of expansions”.

✓ Example 3.6.35 Another difficult limit

In this example we'll evaluate another harder limit, namely

$$\lim_{x \rightarrow 0} \frac{\log\left(\frac{\sin x}{x}\right)}{x^2}$$

The first thing to notice about this limit is that, as x tends to zero, the denominator x^2 tends to 0. So, yet again, to find the limit, we are going to have to show that the numerator also tends to 0 and we are going to have to exhibit a cancellation between the numerator and the denominator.

Because the denominator is x^2 any terms in the numerator, $\log\left(\frac{\sin x}{x}\right)$ that are of order x^3 or higher will contribute terms in the ratio $\frac{\log\left(\frac{\sin x}{x}\right)}{x^2}$ that are of order x or higher. Those terms in the ratio will converge to zero as $x \rightarrow 0$. The moral of this discussion is that we need to compute $\log\left(\frac{\sin x}{x}\right)$ to order x^2 with errors of order x^3 . Now we saw, in Example 3.6.32, that

$$\frac{\sin x}{x} = 1 - \frac{1}{3!}x^2 + O(x^4)$$

We also saw, in equation 3.6.30 with $n = 1$, that

$$\log(1+X) = X + O(X^2)$$

Substituting²⁹ $X = -\frac{1}{3!}x^2 + O(x^4)$, and using that $X^2 = O(x^4)$ (by Remark 3.6.31(b,c)), we have that the numerator

$$\log\left(\frac{\sin x}{x}\right) = \log(1 + X) = X + O(X^2) = -\frac{1}{3!}x^2 + O(x^4)$$

and the limit

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{\log\left(\frac{\sin x}{x}\right)}{x^2} &= \lim_{x \rightarrow 0} \frac{-\frac{1}{3!}x^2 + O(x^4)}{x^2} = \lim_{x \rightarrow 0} \left[-\frac{1}{3!} + O(x^2)\right] = -\frac{1}{3!} \\ &= -\frac{1}{6} \end{aligned}$$

✓ Example 3.6.36 Yet another difficult limit

Evaluate

$$\lim_{x \rightarrow 0} \frac{e^{x^2} - \cos x}{\log(1+x) - \sin x}$$

Solution:

Step 1: Find the limit of the denominator.

$$\lim_{x \rightarrow 0} [\log(1+x) - \sin x] = \log(1+0) - \sin 0 = 0$$

This tells us that we can't evaluate the limit just by finding the limits of the numerator and denominator separately and then dividing.

Step 2: Determine the leading order behaviour of the denominator near $x = 0$. By equations 3.6.30 and 3.6.26,

$$\begin{aligned} \log(1+x) &= x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \dots \\ \sin x &= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots \end{aligned}$$

Taking the difference of these expansions gives

$$\log(1+x) - \sin x = -\frac{1}{2}x^2 + \left(\frac{1}{3} + \frac{1}{3!}\right)x^3 + \dots$$

This tells us that, for x near zero, the denominator is $-\frac{x^2}{2}$ (that's the leading order term) plus contributions that are of order x^3 and smaller. That is

$$\log(1+x) - \sin x = -\frac{x^2}{2} + O(|x|^3)$$

Step 3: Determine the behaviour of the numerator near $x = 0$ to order x^2 with errors of order x^3 and smaller (just like the denominator). By equation 3.6.28

$$e^X = 1 + X + O(X^2)$$

Substituting $X = x^2$

$$\begin{aligned} e^{x^2} &= 1 + x^2 + O(x^4) \\ \cos x &= 1 - \frac{1}{2}x^2 + O(x^4) \end{aligned}$$

by equation 3.6.26. Subtracting, the numerator

$$e^{x^2} - \cos x = \frac{3}{2}x^2 + O(x^4)$$

Step 4: Evaluate the limit.

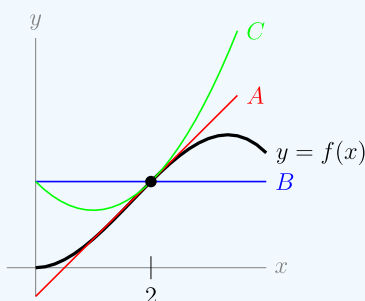
$$\begin{aligned} \lim_{x \rightarrow 0} \frac{e^{x^2} - \cos x}{\log(1+x) - \sin x} &= \lim_{x \rightarrow 0} \frac{\frac{3}{2}x^2 + O(x^4)}{-\frac{x^2}{2} + O(|x|^3)} \\ &= \lim_{x \rightarrow 0} \frac{\frac{3}{2} + O(x^2)}{-\frac{1}{2} + O(|x|)} \\ &= \frac{\frac{3}{2}}{-\frac{1}{2}} = -3 \end{aligned}$$

Exercises

Stage 1

? 1

Below is a graph of $y = f(x)$, along with the constant approximation, linear approximation, and quadratic approximation centered at $a = 2$. Which is which?



? 2

Suppose $T(x)$ is the Taylor series for $f(x) = \arctan^3(e^x + 7)$ centered at $a = 5$. What is $T(5)$?

? 3

Below are a list of common functions, and their Taylor series representations. Match the function to the Taylor series.

function	series
A. $\frac{1}{1-x}$	I. $\sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1}$
B. $\log(1+x)$	II. $\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$
C. $\arctan x$	III. $\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$
D. e^x	IV. $\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$
E. $\sin x$	V. $\sum_{n=0}^{\infty} x^n$
F. $\cos x$	VI. $\sum_{n=0}^{\infty} \frac{x^n}{n!}$

? 4

1. Suppose $f(x) = \sum_{n=0}^{\infty} \frac{n^2}{(n!+1)} (x-3)^n$ for all real x . What is $f^{(20)}(3)$ (the twentieth derivative of $f(x)$ at $x=3$)?
2. Suppose $g(x) = \sum_{n=0}^{\infty} \frac{n^2}{(n!+1)} (x-3)^{2n}$ for all real x . What is $g^{(20)}(3)$?
3. If $h(x) = \frac{\arctan(5x^2)}{x^4}$, what is $h^{(20)}(0)$? What is $h^{(22)}(0)$?

Stage 2

In Questions 5 through 8, you will create Taylor series from scratch. In practice, it is often preferable to modify an existing series, rather than creating a new one, but you should understand both ways.

? 5

Using the definition of a Taylor series, find the Taylor series for $f(x) = \log(x)$ centered at $x=1$.

? 6

Find the Taylor series for $f(x) = \sin x$ centered at $a = \pi$.

? 7

Using the definition of a Taylor series, find the Taylor series for $g(x) = \frac{1}{x}$ centered at $x=10$. What is the interval of convergence of the resulting series?

? 8

Using the definition of a Taylor series, find the Taylor series for $h(x) = e^{3x}$ centered at $x=a$, where a is some constant. What is the radius of convergence of the resulting series?

In Questions 9 through 16, practice creating new Taylor series by modifying known Taylor series, rather than creating your series from scratch.

? 9 (*)

Find the Maclaurin series for $f(x) = \frac{1}{2x-1}$.

? 10 (*)

Let $\sum_{n=0}^{\infty} b_n x^n$ be the Maclaurin series for $f(x) = \frac{3}{x+1} - \frac{1}{2x-1}$,

i.e. $\sum_{n=0}^{\infty} b_n x^n = \frac{3}{x+1} - \frac{1}{2x-1}$.

Find b_n .

? 11 (*)

Find the coefficient c_5 of the fifth degree term in the Maclaurin series $\sum_{n=0}^{\infty} c_n x^n$ for e^{3x} .

? 12 (*)

Express the Taylor series of the function

$$f(x) = \log(1 + 2x)$$

about $x = 0$ in summation notation.

? 13 (*)

The first two terms in the Maclaurin series for $x^2 \sin(x^3)$ are $ax^5 + bx^{11}$, where a and b are constants. Find the values of a and b .

? 14 (*)

Give the first two nonzero terms in the Maclaurin series for $\int \frac{e^{-x^2} - 1}{x} dx$.

? 15 (*)

Find the Maclaurin series for $\int x^4 \arctan(2x) dx$.

? 16 (*)

Suppose that $\frac{df}{dx} = \frac{x}{1 + 3x^3}$ and $f(0) = 1$. Find the Maclaurin series for $f(x)$.

In past chapters, we were only able to exactly evaluate very specific types of series: geometric and telescoping. In Questions 17 through 25, we expand our range by relating given series to Taylor series.

? 17 (*)

The Maclaurin series for $\arctan x$ is given by

$$\arctan x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$$

which has radius of convergence equal to 1. Use this fact to compute the exact value of the series below:

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)3^n}$$

? 18 (*)

Evaluate $\sum_{n=0}^{\infty} \frac{(-1)^n}{n!}$.

? 19 (*)

Evaluate $\sum_{k=0}^{\infty} \frac{1}{e^k k!}$.

? 20 (*)

Evaluate the sum of the convergent series $\sum_{k=1}^{\infty} \frac{1}{\pi^k k!}$.

? 21 (*)

Evaluate $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n 2^n}$.

? 22 (*)

Evaluate $\sum_{n=1}^{\infty} \frac{n+2}{n!} e^n$.

? 23

Evaluate $\sum_{n=1}^{\infty} \frac{2^n}{n}$, or show that it diverges.

? 24

Evaluate

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} \left(\frac{\pi}{4}\right)^{2n+1} (1+2^{2n+1})$$

or show that it diverges.

? 25 (*)

(a) Show that the power series $\sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}$ converges absolutely for all real numbers x .

(b) Evaluate $\sum_{n=0}^{\infty} \frac{1}{(2n)!}$.

? 26

- Using the fact that $\arctan(1) = \frac{\pi}{4}$, how many terms of the Taylor series for arctangent would you have to add up to approximate π with an error of at most 4×10^{-5} ?
- Example 3.6.15 mentions the formula

$$\pi = 16 \arctan \frac{1}{5} - 4 \arctan \frac{1}{239}$$

Using the Taylor series for arctangent, how many terms would you have to add up to approximate π with an error of at most 4×10^{-5} ?

- Assume without proof the following:

$$\arctan \frac{1}{2} + \arctan \frac{1}{3} = \arctan \left(\frac{3+2}{2 \cdot 3 - 1} \right)$$

Using the Taylor series for arctangent, how many terms would you have to add up to approximate π with an error of at most 4×10^{-5} ?

? 27

Suppose you wanted to approximate the number $\log(1.5)$ as a rational number using the Taylor expansion of $\log(1+x)$. How many terms would you need to add to get 10 decimal places of accuracy? (That is, an absolute error less than 5×10^{-11} .)

? 28

Suppose you wanted to approximate the number e as a rational number using the Maclaurin expansion of e^x . How many terms would you need to add to get 10 decimal places of accuracy? (That is, an absolute error less than 5×10^{-11} .)

You may assume without proof that $2 < e < 3$.

? 29

Suppose you wanted to approximate the number $\log(0.9)$ as a rational number using the Taylor expansion of $\log(1-x)$. Which partial sum should you use to get 10 decimal places of accuracy? (That is, an absolute error less than 5×10^{-11} .)

? 30

Define the hyperbolic sine function as

$$\sinh x = \frac{e^x - e^{-x}}{2}.$$

Suppose you wanted to approximate the number $\sinh(b)$ using the Maclaurin series of $\sinh x$, where b is some number in $(-2, 1)$. Which partial sum should you use to guarantee 10 decimal places of accuracy? (That is, an absolute error less than 5×10^{-11} .)

You may assume without proof that $2 < e < 3$.

? 31

Let $f(x)$ be a function with

$$f^{(n)}(x) = \frac{(n-1)!}{2} [(1-x)^{-n} + (-1)^{n-1}(1+x)^{-n}]$$

for all $n \geq 1$.

Give reasonable bounds (both upper and lower) on the error involved in approximating $f(-\frac{1}{3})$ using the partial sum S_6 of the Taylor series for $f(x)$ centered at $a = \frac{1}{2}$.

Remark: One function with this quality is the inverse hyperbolic tangent function³⁰.

Stage 3

? 32 (*)

Use series to evaluate $\lim_{x \rightarrow 0} \frac{1 - \cos x}{1 + x - e^x}$.

? 33 (*)

Evaluate $\lim_{x \rightarrow 0} \frac{\sin x - x + \frac{x^3}{6}}{x^5}$.

? 34

Evaluate $\lim_{x \rightarrow 0} (1 + x + x^2)^{2/x}$ using a Taylor series for the natural logarithm.

? 35

Use series to evaluate

$$\lim_{x \rightarrow \infty} \left(1 + \frac{1}{2x}\right)^x$$

? 36

Evaluate the series $\sum_{n=0}^{\infty} \frac{(n+1)(n+2)}{7^n}$ or show that it diverges.

? 37

Write the series $f(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+4}}{(2n+1)(2n+2)}$ as a combination of familiar functions.

? 38

1. Find the Maclaurin series for $f(x) = (1-x)^{-1/2}$. What is its radius of convergence?
2. Manipulate the series you just found to find the Maclaurin series for $g(x) = \arcsin x$. What is its radius of convergence?

? 39 (*)

Find the Taylor series for $f(x) = \log(x)$ centered at $a = 2$. Find the interval of convergence for this series.

? 40 (*)

$$\text{Let } I(x) = \int_0^x \frac{1}{1+t^4} dt.$$

1. Find the Maclaurin series for $I(x)$.
2. Approximate $I(1/2)$ to within ± 0.0001 .
3. Is your approximation in (b) larger or smaller than the true value of $I(1/2)$? Explain.

? 41 (*)

Using a Maclaurin series, the number $a = 1/5 - 1/7 + 1/18$ is found to be an approximation for $I = \int_0^1 x^4 e^{-x^2} dx$. Give the best upper bound you can for $|I - a|$.

? 42 (*)

Find an interval of length 0.0002 or less that contains the number

$$I = \int_0^{\frac{1}{2}} x^2 e^{-x^2} dx$$

? 43 (*)

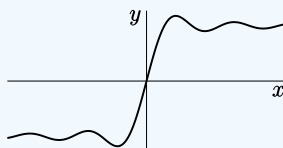
Let $I(x) = \int_0^x \frac{e^{-t} - 1}{t} dt$.

1. Find the Maclaurin series for $I(x)$.
2. Approximate $I(1)$ to within ± 0.01 .
3. Explain why your answer to part (b) has the desired accuracy.

? 44 (*)

The function $\Sigma(x)$ is defined by $\Sigma(x) = \int_0^x \frac{\sin t}{t} dt$.

1. Find the Maclaurin series for $\Sigma(x)$.
2. It can be shown that $\Sigma(x)$ has an absolute maximum which occurs at its smallest positive critical point (see the graph of $\Sigma(x)$ below). Find this critical point.
3. Use the previous information to find the maximum value of $\Sigma(x)$ to within ± 0.01 .



? 45 (*)

Let $I(x) = \int_0^x \frac{\cos t - 1}{t^2} dt$.

1. Find the Maclaurin series for $I(x)$.
2. Use this series to approximate $I(1)$ to within ± 0.01 .
3. Is your estimate in (b) greater than $I(1)$? Explain.

? 46 (*)

Let $I(x) = \int_0^x \frac{\cos t + t \sin t - 1}{t^2} dt$

1. Find the Maclaurin series for $I(x)$.
2. Use this series to approximate $I(1)$ to within ± 0.001 .
3. Is your estimate in (b) greater than or less than $I(1)$?

? 47 (*)

Define $f(x) = \int_0^x \frac{1 - e^{-t}}{t} dt$.

1. Show that the Maclaurin series for $f(x)$ is $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n \cdot n!} x^n$.
2. Use the ratio test to determine the values of x for which the Maclaurin series $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n \cdot n!} x^n$ converges.

? 48 (*)

Show that $\int_0^1 \frac{x^3}{e^x - 1} dx \leq \frac{1}{3}$.

? 49 (*)

Let $\cosh(x) = \frac{e^x + e^{-x}}{2}$.

1. Find the power series expansion of $\cosh(x)$ about $x_0 = 0$ and determine its interval of convergence.
2. Show that $3\frac{2}{3} \leq \cosh(2) \leq 3\frac{2}{3} + 0.1$.
3. Show that $\cosh(t) \leq e^{\frac{1}{2}t^2}$ for all t .

? 50

The law of the instrument says “If you have a hammer then everything looks like a nail” — it is really a description of the “tendency of jobs to be adapted to tools rather than adapting tools to jobs”³¹. Anyway, this is a long way of saying that just because we know how to compute things using Taylor series doesn't mean we should neglect other techniques.

1. Using Newton's method, approximate the constant $\sqrt[3]{2}$ as a root of the function $g(x) = x^3 - 2$. Using a calculator, make your estimation accurate to within 0.01.
2. You may assume without proof that

$$\sqrt[3]{x} = 1 + \frac{1}{6}(x-1) + \sum_{n=2}^{\infty} (-1)^{n-1} \frac{(2)(5)(8) \cdots (3n-4)}{3^n n!} (x-1)^n.$$

for all real numbers x . Using the fact that this is an alternating series, how many terms would you have to add for the partial sum to estimate $\sqrt[3]{2}$ with an error less than 0.01?

? 51

Let $f(x) = \arctan(x^3)$. Write $f^{(10)}\left(\frac{1}{5}\right)$ as a sum of rational numbers with an error less than 10^{-6} using the Maclaurin series for arctangent.

? 52

Consider the following function:

$$f(x) = \begin{cases} e^{-1/x^2} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

1. Sketch $y = f(x)$.
2. Assume (without proof) that $f^{(n)}(0) = 0$ for all whole numbers n . Find the Maclaurin series for $f(x)$.
3. Where does the Maclaurin series for $f(x)$ converge?
4. For which values of x is $f(x)$ equal to its Maclaurin series?

? 53

Suppose $f(x)$ is an odd function, and $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$. Simplify $\sum_{n=0}^{\infty} \frac{f^{(2n)}(0)}{(2n)!} x^{2n}$.

1. Please review your notes from last term if this material is feeling a little unfamiliar.
2. Did you take a quick look at your notes?

3. This is probably the most commonly used formula for the error. But there is another fairly commonly used formula. It, and some less commonly used formulae, are given in the next (optional) subsection “More about the Taylor Remainder”.
4. The discussion here is only supposed to jog your memory. If it is feeling insufficiently jogged, then please look at your notes from last term.
5. The reader might ask whether or not we will give the series for other trigonometric functions or their inverses. While the tangent function has a perfectly well defined series, its coefficients are not as simple as those of the series we have seen — they form a sequence of numbers known (perhaps unsurprisingly) as the “tangent numbers”. They, and the related Bernoulli numbers, have many interesting properties, links to which the interested reader can find with their favourite search engine. The Maclaurin series for inverse sine is $\arcsin(x) = \sum_{n=0}^{\infty} \frac{4^{-n}}{2n+1} \frac{(2n)!}{(n!)^2} x^{2n+1}$ which is quite tidy, but proving it is beyond the scope of the course.
6. Warning: antique sign–sine pun. No doubt the reader first saw it many years syne.
7. Recall, you studied that differential equation in the section on separable differential equations (Theorem 2.4.4 in Section 2.4) as well as wayyyy back in the section on exponential growth and decay in differential calculus.
8. Recall that when we solve of a separable differential equation our general solution will have an arbitrary constant in it. That constant cannot be determined from the differential equation alone and we need some extra data to find it. This extra information is often information about the system at its beginning (for example when position or time is zero) — hence “initial conditions”. Of course the reader is already familiar with this because it was covered back in Section 2.4.
9. While the use of the ideas of induction goes back over 2000 years, the first recorded rigorous use of induction appeared in the work of Levi ben Gershon (1288–1344, better known as Gersonides). The first explicit formulation of mathematical induction was given by the French mathematician Blaise Pascal in 1665.
10. In Theorem 3.4.38 in the CLP-1 text, we assumed, for simplicity, that $a < b$. To get (GVMT) when $b < a$ simply exchange a and b in Theorem 3.4.38.
11. Note that the function G need not be related to f . It just has to be differentiable with a nonzero derivative.
12. The computation of π has a very, very long history and your favourite search engine will turn up many sites that explore the topic. For a more comprehensive history one can turn to books such as “A history of Pi” by Petr Beckmann and “The joy of π ” by David Blatner.
13. If you don't know what this means (forgive the pun) don't worry, because it is not part of the course. Standard deviation a way of quantifying variation within a population.
14. We could get a computer algebra system to do it for us without much difficulty — but we wouldn't learn much in the process. The point of this example is to illustrate that one can do more than just represent a function with Taylor series. More on this in the next section.
15. Check the derivative!
16. The authors hope that by now we all “know” that e is between 2 and 3, but maybe we don't know how to prove it.
17. We do not wish to give a primer on imaginary and complex numbers here. The interested reader can start by looking at Appendix B.
18. It is worth mentioning here that history of this topic is perhaps a little rough on Roger Cotes (1682–1716) who was one of the strongest mathematicians of his time and a collaborator of Newton. Cotes published a paper on logarithms in 1714 in which he states $i x = \log(\cos x + i \sin x)$. (after translating his results into more modern notation). He proved this result by computing in two different ways the surface area of an ellipse rotated about one axis and equating the results. Unfortunately Cotes died only 2 years later at the age of 33. Upon hearing of his death Newton is supposed to have said “If he had lived, we might have known something.” The reader might think this a rather weak statement, however coming from Newton it was high praise.
19. We are hiding some mathematics behind this “consequently”. What we are really using our knowledge of Taylor polynomials to write $f(x) = \sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + E_5(x)$ where $E_5(x) = \frac{f^{(6)}(c)}{6!}x^6$ and c is between 0 and x . We are effectively hiding “ $E_5(x)$ ” inside the “.....”. Now we can divide both sides by x (assuming $x \neq 0$): $\frac{\sin(x)}{x} = 1 - \frac{1}{3!}x^2 + \frac{1}{5!}x^4 + \frac{E_5(x)}{x}$. and everything is fine provided the term $\frac{E_5(x)}{x}$ stays well behaved.
20. Many of you learned about l'Hôpital's rule in school and all of you should have seen it last term in your differential calculus course.
21. It takes 3 applications of l'Hôpital's rule and some careful cleaning up of the intermediate expressions. Oof!
22. Though there were a few comments in a footnote.
23. It is worth pointing out that our Taylor series must be expanded about the point to which we are limiting — i.e. a . To work out a limit as $x \rightarrow a$ we need Taylor series expanded about a and not some other point.

24. To be precise, C and D do not depend on x , though they may, and usually do, depend on m .
25. Remember that $n! = 1 \times 2 \times 3 \times \cdots \times n$, and that we use the convention $0! = 1$.
26. It is not too hard to make this rigorous using the principle of mathematical induction. The interested reader should do a little search-engine-ing. Induction is a very standard technique for proving statements of the form “For every natural number n, \dots ”. For example For every natural number n , $\sum_{k=1}^n k = \frac{n(n+1)}{2}$, or
For every natural number n , $\frac{d^n}{dx^n} \{\log(1+x)\} = (-1)^{n-1} \frac{(n-1)!}{(1+x)^n}$. It was also used by Polya (1887–1985) to give a very convincing (but subtly (and deliberately) flawed) proof that all horses have the same colour.
27. Since $|c| \leq \frac{1}{2}$, $-\frac{1}{2} \leq c \leq \frac{1}{2}$. If we now add 1 to every term we get $\frac{1}{2} \leq 1+c \leq \frac{3}{2}$ and so $|1+c| \geq \frac{1}{2}$. You can also do this with the triangle inequality which tells us that for any x, y we know that $|x+y| \leq |x|+|y|$. Actually, you want the reverse triangle inequality (which is a simple corollary of the triangle inequality) which says that for any x, y we have $|x+y| \geq ||x|-|y||$.
28. Use of l'Hôpital's rule here could be characterised as a “courageous decision”. The interested reader should search-engine their way to Sir Humphrey Appleby and “Yes Minister” to better understand this reference (and the workings of government in the Westminster system). Discretion being the better part of valour, we'll stop and think a little before limiting (ha) our choices.
29. In our derivation of $\log(1+X) = X + O(X^2)$ in Example 3.6.29, we required only that $|X| \leq \frac{1}{2}$. So we are free to substitute $X = -\frac{1}{3!}x^2 + O(x^4)$ for any x that is small enough that $|\frac{1}{3!}x^2 + O(x^4)| < \frac{1}{2}$.
30. Of course it is! Actually, hyperbolic tangent is $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, and inverse hyperbolic tangent is its functional inverse.
31. Quote from Silvan Tomkins's *Computer Simulation of Personality: Frontier of Psychological Theory*. See also Birmingham screwdrivers.

This page titled [3.6: Taylor Series](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

3.7: Optional — Rational and irrational numbers

In this optional section we shall use series techniques to look a little at rationality and irrationality of real numbers. We shall see the following results.

- A real number is rational (i.e. a ratio of two integers) if and only if its decimal expansion is eventually periodic. “Eventually periodic” means that, if we denote the n^{th} decimal place by d_n , then there are two positive integers k and p such that $d_{n+p} = d_n$ whenever $n > k$. So the part of the decimal expansion after the decimal point looks like

$$\cdot \underbrace{a_1 a_2 a_3 \cdots a_k}_{\text{non-repeating}} \underbrace{b_1 b_2 \cdots b_p}_{\text{repeating}} \underbrace{b_1 b_2 \cdots b_p}_{\text{repeating}} \underbrace{b_1 b_2 \cdots b_p}_{\text{repeating}} \cdots$$

It is possible that a finite number of decimal places right after the decimal point do not participate in the periodicity. It is also possible that $p = 1$ and $b_1 = 0$, so that the decimal expansion ends with an infinite string of zeros.

- e is irrational.
- π is irrational.

Decimal expansions of rational numbers

We start by showing that a real number is rational if and only if its decimal expansion is eventually periodic. We need only consider the expansions of numbers $0 < x < 1$. If a number is negative then we can just multiply it by -1 and not change the expansion. Similarly if the number is larger than 1 then we can just subtract off the integer part of the number and leave the expansion unchanged.

Eventually periodic implies rational

Let us assume that a number $0 < x < 1$ has a decimal expansion that is eventually periodic. Hence we can write

$$x = 0.\underbrace{a_1 a_2 a_3 \cdots a_k}_{\text{non-repeating}} \underbrace{b_1 b_2 \cdots b_p}_{\text{repeating}} \underbrace{b_1 b_2 \cdots b_p}_{\text{repeating}} \underbrace{b_1 b_2 \cdots b_p}_{\text{repeating}} \cdots$$

Let $\alpha = a_1 a_2 a_3 \cdots a_k$ and $\beta = b_1 b_2 \cdots b_p$. In particular, α has at most k digits and β has at most p digits. Then we can (carefully) write

$$\begin{aligned} x &= \frac{\alpha}{10^k} + \frac{\beta}{10^{k+p}} + \frac{\beta}{10^{k+2p}} + \frac{\beta}{10^{k+3p}} + \cdots \\ &= \frac{\alpha}{10^k} + \frac{\beta}{10^{k+p}} \sum_{j=0}^{\infty} 10^{-jp} \end{aligned}$$

This sum is just a geometric series (see Lemma 3.2.5) and we can evaluate it

$$\begin{aligned} &= \frac{\alpha}{10^k} + \frac{\beta}{10^{k+p}} \cdot \frac{1}{1 - 10^{-p}} = \frac{\alpha}{10^k} + \frac{\beta}{10^k} \cdot \frac{1}{10^p - 1} \\ &= \frac{1}{10^k} \left(\alpha + \frac{\beta}{10^p - 1} \right) = \frac{\alpha(10^p - 1) + \beta}{10^k(10^p - 1)} \end{aligned}$$

This is a ratio of integers, so x is a rational number.

Rational implies eventually periodic

Let $0 < x < 1$ be rational with $x = \frac{a}{b}$, where a and b are positive integers. We wish to show that x 's decimal expansion is eventually periodic. Start by looking at the last formula we derived in the “eventually periodic implies rational” subsection. If we can express the denominator b in the form $\frac{10^k(10^p - 1)}{q}$ with k , p and q integers, we will be in business because $\frac{a}{b} = \frac{aq}{10^k(10^p - 1)}$. From this we can generate the desired decimal expansion by running the argument of the last subsection backwards. So we want to find integers k , p , q such that $10^{k+p} - 10^k = b \cdot q$. To do so consider the powers of 10 up to 10^b :

$$1, 10^1, 10^2, 10^3, \dots, 10^b$$

For each $j = 0, 1, 2, \dots, b$, find integers c_j and $0 \leq r_j < b$ so that

$$10^j = b \cdot c_j + r_j$$

To do so, start with 10^j and repeatedly subtract b from it until the remainder drops strictly below b . The r_j 's can take at most b different values, namely $0, 1, 2, \dots, b-1$, and we now have $b+1$ r_j 's, namely r_0, r_1, \dots, r_b . So we must be able to find two powers of 10 which give the same remainder¹. That is there must be $0 \leq k < l \leq b$ so that $r_k = r_l$. Hence

$$\begin{aligned} 10^l - 10^k &= (bc_l + r_l) - (bc_k + r_k) \\ &= b(c_l - c_k) \qquad \text{since } r_k = r_l. \end{aligned}$$

and we have

$$b = \frac{10^k(10^p - 1)}{q}$$

where $p = l - k$ and $q = c_l - c_k$ are both strictly positive integers, since $l > k$ so that $10^l - 10^k > 0$. Thus we can write

$$\frac{a}{b} = \frac{aq}{10^k(10^p - 1)}$$

Next divide the numerator aq by $10^p - 1$ and compute the remainder. That is, write $aq = \alpha(10^p - 1) + \beta$ with $0 \leq \beta < 10^p - 1$. Notice that $0 \leq \alpha < 10^k$, as otherwise $x = \frac{a}{b} \geq 1$. That is, α has at most k digits and β has at most p digits. This, finally, gives us

$$\begin{aligned} x &= \frac{a}{b} = \frac{\alpha(10^p - 1) + \beta}{10^k(10^p - 1)} \\ &= \frac{\alpha}{10^k} + \frac{\beta}{10^k(10^p - 1)} \\ &= \frac{\alpha}{10^k} + \frac{\beta}{10^{k+p}(1 - 10^{-p})} \\ &= \frac{\alpha}{10^k} + \frac{\beta}{10^{k+p}} \sum_{j=0}^{\infty} 10^{-pj} \end{aligned}$$

which gives the required eventually periodic expansion.

Irrationality of e

We will give 2 proofs that the number e is irrational, the first due to Fourier (1768–1830) and the second due to Pennisi (1918–2010). Both are proofs by contradiction² — we first assume that e is rational and then show that this implies a contradiction. In both cases we reach the contradiction by showing that a given quantity (related to the series expression for e) must be both a positive integer and also strictly less than 1.

Proof 1

This proof is due to Fourier. Let us assume that the number e is rational so we can write it as

$$e = \frac{a}{b}$$

where a, b are positive integers. Using the Maclaurin series for e^x we have

$$\frac{a}{b} = e^1 = \sum_{n=0}^{\infty} \frac{1}{n!}$$

Now multiply both sides by $b!$ to get

$$a \frac{b!}{b} = \sum_{n=0}^{\infty} \frac{b!}{n!}$$

The left-hand side of this expression is an integer. We complete the proof by showing that the right-hand side cannot be an integer (and hence that we have a contradiction).

First split the series on the right-hand side into two pieces as follows

$$\sum_{n=0}^{\infty} \frac{b!}{n!} = \underbrace{\sum_{n=0}^b \frac{b!}{n!}}_{=A} + \underbrace{\sum_{n=b+1}^{\infty} \frac{b!}{n!}}_{=B}$$

The first sum, A , is finite sum of integers:

$$A = \sum_{n=0}^b \frac{b!}{n!} = \sum_{n=0}^b (n+1)(n+2)\cdots(b-1)b.$$

Consequently A must be an integer. Notice that we simplified the ratio of factorials using the fact that when $b \geq n$ we have

$$\frac{b!}{n!} = \frac{1 \cdot 2 \cdots n(n+1)(n+2)\cdots(b-1)b}{1 \cdot 2 \cdots n} = (n+1)(n+2)\cdots(b-1)b.$$

Now we turn to the second sum. Since it is a sum of strictly positive terms we must have

$$B > 0$$

We complete the proof by showing that $B < 1$. To do this we bound each term from above:

$$\begin{aligned} \frac{b!}{n!} &= \frac{1}{\underbrace{(b+1)(b+2)\cdots(n-1)n}_{n-b \text{ factors}}} \\ &\leq \frac{1}{\underbrace{(b+1)(b+1)\cdots(b+1)(b+1)}_{n-b \text{ factors}}} = \frac{1}{(b+1)^{n-b}} \end{aligned}$$

Indeed the inequality is strict except when $n = b + 1$. Hence we have that

$$\begin{aligned} B &< \sum_{n=b+1}^{\infty} \frac{1}{(b+1)^{n-b}} \\ &= \frac{1}{(b+1)} + \frac{1}{(b+1)^2} + \frac{1}{(b+1)^3} + \cdots \end{aligned}$$

This is just a geometric series (see Lemma 3.2.5) and equals

$$\begin{aligned} &= \frac{1}{(b+1)} \frac{1}{1 - \frac{1}{b+1}} \\ &= \frac{1}{b+1-1} = \frac{1}{b} \end{aligned}$$

And since b is a positive integer, we have shown that

$$0 < B < 1$$

and thus B cannot be an integer.

Thus we have that

$$\underbrace{\frac{b!}{b}}_{\text{integer}} = \underbrace{A}_{\text{integer}} + \underbrace{B}_{\text{not integer}}$$

which gives a contradiction. Thus e cannot be rational.

Proof 2

This proof is due to Pennisi (1953). Let us (again) assume that the number e is rational. Hence it can be written as

$$e = \frac{a}{b},$$

where a, b are positive integers. This means that we can write

$$e^{-1} = \frac{b}{a}.$$

Using the Maclaurin series for e^x we have

$$\frac{b}{a} = e^{-1} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}$$

Before we do anything else, we multiply both sides by $(-1)^{a+1}a!$ — this might seem a little strange at this point, but the reason will become clear as we proceed through the proof. The expression is now

$$(-1)^{a+1}b\frac{a!}{a} = \sum_{n=0}^{\infty} \frac{(-1)^{n+a+1}a!}{n!}$$

The left-hand side of the expression is an integer. We again complete the proof by showing that the right-hand side cannot be an integer.

We split the series on the right-hand side into two pieces:

$$\sum_{n=0}^{\infty} \frac{(-1)^{n+a+1}a!}{n!} = \underbrace{\sum_{n=0}^a \frac{(-1)^{n+a+1}a!}{n!}}_{=A} + \underbrace{\sum_{n=a+1}^{\infty} \frac{(-1)^{n+a+1}a!}{n!}}_{=B}$$

We will show that A is an integer while $0 < B < 1$; this gives the required contradiction.

Every term in the sum A is an integer. To see this we simplify the ratio of factorials as we did in the previous proof:

$$A = \sum_{n=0}^a \frac{(-1)^{n+a+1}a!}{n!} = \sum_{n=0}^a (-1)^{n+a+1} (n+1)(n+2)\cdots(a-1)a$$

Let us now examine the series B . Again clean up the ratio of factorials:

$$\begin{aligned} B &= \sum_{n=a+1}^{\infty} \frac{(-1)^{n+a+1}a!}{n!} = \sum_{n=a+1}^{\infty} \frac{(-1)^{n+a+1}}{(a+1)\cdot(a+2)\cdots(n-1)\cdot n} \\ &= \frac{(-1)^{2a+2}}{a+1} + \frac{(-1)^{2a+3}}{(a+1)(a+2)} + \frac{(-1)^{2a+4}}{(a+1)(a+2)(a+3)} + \cdots \\ &= \frac{1}{a+1} - \frac{1}{(a+1)(a+2)} + \frac{1}{(a+1)(a+2)(a+3)} - \cdots \end{aligned}$$

Hence B is an alternating series of decreasing terms and by the alternating series test (Theorem 3.3.14) it converges. Further, it must converge to a number between its first and second partial sums (see the discussion before Theorem 3.3.14). Hence the right-hand side lies between

$$\frac{1}{a+1} \quad \text{and} \quad \frac{1}{a+1} - \frac{1}{(a+1)(a+2)} = \frac{1}{a+2}$$

Since a is a positive integer the above tells us that B converges to a real number strictly greater than 0 and strictly less than 1. Hence it cannot be an integer.

This gives us a contradiction and hence e cannot be rational.

Irrationality of π

This proof is due to Niven (1946) and doesn't require any mathematics beyond the level of this course. Much like the proofs above we will start by assuming that π is rational and then reach a contradiction. Again this contradiction will be that a given quantity must be an integer but at the same time must lie strictly between 0 and 1.

Assume that π is a rational number and so can be written as $\pi = \frac{a}{b}$ with a, b positive integers. Now let n be a positive integer and define the polynomial

$$f(x) = \frac{x^n(a - bx)^n}{n!}.$$

It is certainly not immediately obvious why and how Niven chose this polynomial, but you will see that it has been very carefully crafted to make the proof work. In particular we will show — under our assumption that π is rational — that, if n is really big, then

$$I_n = \int_0^\pi f(x) \sin(x) dx$$

is an integer and it also lies strictly between 0 and 1, giving the required contradiction.

Bounding the integral

Consider again the polynomial

$$f(x) = \frac{x^n(a - bx)^n}{n!}.$$

Notice that

$$\begin{aligned} f(0) &= 0 \\ f(\pi) &= f(a/b) = 0. \end{aligned}$$

Furthermore, for $0 \leq x \leq \pi = a/b$, we have $x \leq \frac{a}{b}$ and $a - bx \leq a$ so that

$$0 \leq x(a - bx) \leq a^2/b.$$

We could work out a more precise³ upper bound, but this one is sufficient for the analysis that follows. Hence

$$0 \leq f(x) \leq \left(\frac{a^2}{b}\right)^n \frac{1}{n!}$$

We also know that for $0 \leq x \leq \pi = a/b$, $0 \leq \sin(x) \leq 1$. Thus

$$0 \leq f(x) \sin(x) \leq \left(\frac{a^2}{b}\right)^n \frac{1}{n!}$$

for all $0 \leq x \leq 1$. Using this inequality we bound

$$0 < I_n = \int_0^\pi f(x) \sin(x) dx < \left(\frac{a^2}{b}\right)^n \frac{1}{n!}.$$

We will later show that, if n is really big, then $\left(\frac{a^2}{b}\right)^n \frac{1}{n!} < 1$. We'll first show, starting now, that I_n is an integer.

Integration by parts

In order to show that the value of this integral is an integer we will use integration by parts. You have already practiced using integration by parts to integrate quantities like

$$\int x^2 \sin(x) dx$$

and this integral isn't much different. For the moment let us just use the fact that $f(x)$ is a polynomial of degree $2n$. Using integration by parts with $u = f(x)$, $dv = \sin(x)$ and $v = -\cos(x)$ gives us

$$\int f(x) \sin(x) dx = -f(x) \cos(x) + \int f'(x) \cos(x) dx$$

Use integration by parts again with $u = f'(x)$, $dv = \cos(x)$ and $v = \sin(x)$.

$$= -f(x) \cos(x) + f'(x) \sin(x) - \int f''(x) \sin(x) dx$$

Use integration by parts yet again, with $u = f''(x)$, $dv = \sin(x)$ and $v = -\cos(x)$.

$$= -f(x) \cos(x) + f'(x) \sin(x) + f''(x) \cos(x) - \int f'''(x) \cos(x) dx$$

And now we can see the pattern; we get alternating signs, and then derivatives multiplied by sines and cosines:

$$\int f(x) \sin(x) dx = \cos(x) \left(-f(x) + f''(x) - f^{(4)}(x) + f^{(6)}(x) - \dots \right) \\ + \sin(x) \left(f'(x) - f'''(x) + f^{(5)}(x) - f^{(7)}(x) + \dots \right)$$

This terminates at the $2n^{\text{th}}$ derivative since $f(x)$ is a polynomial of degree $2n$. We can check this computation by differentiating the terms on the right-hand side:

$$\frac{d}{dx} \left(\cos(x) \left(-f(x) + f''(x) - f^{(4)}(x) + f^{(6)}(x) - \dots \right) \right) \\ = -\sin(x) \left(-f(x) + f''(x) - f^{(4)}(x) + f^{(6)}(x) - \dots \right) \\ + \cos(x) \left(-f'(x) + f'''(x) - f^{(5)}(x) + f^{(7)}(x) - \dots \right)$$

and similarly

$$\frac{d}{dx} \left(\sin(x) \left(f'(x) - f'''(x) + f^{(5)}(x) - f^{(7)}(x) + \dots \right) \right) \\ = \cos(x) \left(f'(x) - f'''(x) + f^{(5)}(x) - f^{(7)}(x) + \dots \right) \\ + \sin(x) \left(f''(x) - f^{(4)}(x) + f^{(6)}(x) - \dots \right)$$

When we add these two expressions together all the terms cancel except $f(x) \sin(x)$, as required.

Now when we take the definite integral from 0 to π , all the sine terms give 0 because $\sin(0) = \sin(\pi) = 0$. Since $\cos(\pi) = -1$ and $\cos(0) = +1$, we are just left with:

$$\int_0^\pi f(x) \sin(x) dx = \left(f(0) - f''(0) + f^{(4)}(0) - f^{(6)}(0) + \dots + (-1)^n f^{(2n)}(0) \right) \\ + \left(f(\pi) - f''(\pi) + f^{(4)}(\pi) - f^{(6)}(\pi) + \dots + (-1)^n f^{(2n)}(\pi) \right)$$

So to show that I_n is an integer, it now suffices to show that $f^{(j)}(0)$ and $f^{(j)}(\pi)$ are integers.

The derivatives are integers

Recall that

$$f(x) = \frac{x^n (a - bx)^n}{n!}$$

and expand it:

$$f(x) = \frac{c_0}{n!} x^0 + \frac{c_1}{n!} x^1 + \dots + \frac{c_n}{n!} x^n + \dots + \frac{c_{2n}}{n!} x^{2n}$$

All the c_j are integers, and clearly $c_j = 0$ for all $j = 0, 1, \dots, n-1$, because of the factor x^n in $f(x)$.

Now take the k^{th} derivative and set $x = 0$. Note that, if $j < k$, then $\frac{d^k}{dx^k} x^j = 0$ for all x and, if $j > k$, then $\frac{d^k}{dx^k} x^j$ is some number times x^{j-k} which evaluates to zero when we set $x = 0$. So

$$f^{(k)}(0) = \frac{d^k}{dx^k} \left(\frac{c_k}{k!} x^k \right) = \frac{k! c_k}{n!}$$

If $k < n$, then this is zero since $c_k = 0$. If $k > n$, this is an integer because c_k is an integer and $k!/n! = (n+1)(n+2) \dots (k-1)k$ is an integer. If $k = n$, then $f^{(k)}(0) = c_n$ is again an integer. Thus all the derivatives of $f(x)$ evaluated at $x = 0$ are integers.

But what about the derivatives at $\pi = a/b$? To see this, we can make use of a handy symmetry. Notice that

$$f(x) = f(\pi - x) = f(a/b - x)$$

You can confirm this by just grinding through the algebra:

$$\begin{aligned}
 f(x) &= \frac{x^n (a - bx)^n}{n!} && \text{now replace } x \text{ with } a/b - x \\
 f(a/b - x) &= \frac{(a/b - x)^n (a - b(a/b - x))^n}{n!} && \text{start cleaning this up:} \\
 &= \frac{\left(\frac{a-bx}{b}\right)^n (a - a + bx)^n}{n!} \\
 &= \frac{\left(\frac{a-bx}{b}\right)^n (bx)^n}{n!} \\
 &= \frac{(a - bx)^n x^n}{n!} = f(x)
 \end{aligned}$$

Using this symmetry (and the chain rule) we see that

$$f'(x) = -f'(\pi - x)$$

and if we keep differentiating

$$f^{(k)}(x) = (-1)^k f^{(k)}(\pi - x)$$

Setting $x = 0$ in this tells us that

$$f^{(k)}(0) = (-1)^k f^{(k)}(\pi)$$

So because all the derivatives at $x = 0$ are integers, we know that all the derivatives at $x = \pi$ are also integers.

Hence the integral we are interested in

$$\int_0^\pi f(x) \sin(x) dx$$

must be an integer.

Putting it together

Based on our assumption that $\pi = a/b$ is rational, we have shown that the integral

$$I_n = \int_0^\pi \frac{x^n (a - bx)^n}{n!} \sin(x) dx$$

satisfies

$$0 < I_n < \left(\frac{a^2}{b}\right)^n \frac{1}{n!}$$

and also that I_n is an integer.

We are, however, free to choose n to be any positive integer we want. If we take n to be very large — in particular much much larger than a — then $n!$ will be much much larger than a^{2n} (we showed this in Example 3.6.8), and consequently

$$0 < I_n < \left(\frac{a^2}{b}\right)^n \frac{1}{n!} < 1$$

Which means that the integral cannot be an integer. This gives the required contradiction, showing that π is irrational.

1. This is an application of the pigeon hole principle — the very simple but surprisingly useful idea that if you have n items which you have to put in m boxes, and if $n > m$, then at least one box must contain more than one item.

2. Proof by contradiction is a standard and very powerful method of proof in mathematics. It relies on the law of the excluded middle which states that any given mathematical statement P is either true or false. Because of this, if we can show that the statement P being false implies something contradictory — like $1 = 0$ or $a > a$ — then we can conclude that P must be true. The interested reader can certainly find many examples (and a far more detailed explanation) using their favourite search engine.
3. You got lots of practice finding the maximum and minimum values of continuous functions on closed intervals when you took calculus last term.

This page titled [3.7: Optional — Rational and irrational numbers](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

Appendices

A: High School Material

A.1: Similar Triangles

A.2: Pythagoras

A.3: Trigonometry — Definitions

A.4: Radians, Arcs and Sectors

A.5: Trigonometry — Graphs

A.6: Trigonometry — Special Triangles

A.7: Trigonometry — Simple Identities

A.8: Trigonometry — Add and Subtract Angles

A.9: Inverse Trigonometric Functions

A.10: Areas

A.11: Volumes

A.12: Powers

A.13: Logarithms

A.14: Highschool Material You Should be Able to Derive

A.15: Cartesian Coordinates

A.16: Roots of Polynomials

B: Complex Numbers and Exponentials

B.1: Definition and Basic Operations

B.2: The Complex Exponential

C: More About Numerical Integration

C.1: Richardson Extrapolation

C.2: Romberg Integration

C.3: Adaptive Quadrature

D: Numerical Solution of ODE's

D.1: Simple ODE Solvers — Derivation

D.2: Simple ODE Solvers — Error Behaviour

D.3: Variable Step Size Methods

E: Hints for Exercises

F: Answers to Exercises

G: Solutions to Exercises

This page titled [Appendices](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A: High School Material

This chapter is really split into three parts.

- Sections A.1 to A.11 contains results that we expect you to understand and know.
- Then Section A.14 contains results that we don't expect you to memorise, but that we think you should be able to quickly derive from other results you know.
- The remaining sections contain some material (that may be new to you) that is related to topics covered in the main body of these notes.

This page titled [A: High School Material](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.1: Similar Triangles

Two triangles T_1, T_2 are similar when

- (AAA — angle angle angle) The angles of T_1 are the same as the angles of T_2 .
- (SSS — side side side) The ratios of the side lengths are the same. That is

$$\frac{A}{a} = \frac{B}{b} = \frac{C}{c}$$


- (SAS — side angle side) Two sides have lengths in the same ratio and the angle between them is the same. For example

$$\frac{A}{a} = \frac{C}{c} \text{ and angle } \beta \text{ is same}$$

This page titled [A.1: Similar Triangles](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.2: Pythagoras

For a right-angled triangle the length of the hypotenuse is related to the lengths of the other two sides by


$$(\text{adjacent})^2 + (\text{opposite})^2 = (\text{hypotenuse})^2$$

This page titled [A.2: Pythagoras](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.3: Trigonometry — Definitions

$$\begin{aligned}\sin \theta &= \frac{\text{opposite}}{\text{hypotenuse}} & \csc \theta &= \frac{1}{\sin \theta} \\ \cos \theta &= \frac{\text{adjacent}}{\text{hypotenuse}} & \sec \theta &= \frac{1}{\cos \theta} \\ \tan \theta &= \frac{\text{opposite}}{\text{adjacent}} & \cot \theta &= \frac{1}{\tan \theta}\end{aligned}$$

This page titled A.3: Trigonometry — Definitions is shared under a [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.4: Radians, Arcs and Sectors

For a circle of radius r and angle of θ radians:

- Arc length $L(\theta) = r\theta$.
- Area of sector $A(\theta) = \frac{\theta}{2}r^2$.

This page titled [A.4: Radians, Arcs and Sectors](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.5: Trigonometry — Graphs



$\sin \theta$



$\cos \theta$



$\tan \theta$

This page titled A.5: Trigonometry — Graphs is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.6: Trigonometry — Special Triangles

From the above pair of special triangles we have

$$\begin{array}{lll} \sin \frac{\pi}{4} = \frac{1}{\sqrt{2}} & \sin \frac{\pi}{6} = \frac{1}{2} & \sin \frac{\pi}{3} = \frac{\sqrt{3}}{2} \\ \cos \frac{\pi}{4} = \frac{1}{\sqrt{2}} & \cos \frac{\pi}{6} = \frac{\sqrt{3}}{2} & \cos \frac{\pi}{3} = \frac{1}{2} \\ \tan \frac{\pi}{4} = 1 & \tan \frac{\pi}{6} = \frac{1}{\sqrt{3}} & \tan \frac{\pi}{3} = \sqrt{3} \end{array}$$

This page titled A.6: Trigonometry — Special Triangles is shared under a [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.7: Trigonometry — Simple Identities

- Periodicity

$$\sin(\theta + 2\pi) = \sin(\theta) \quad \cos(\theta + 2\pi) = \cos(\theta)$$

- Reflection

$$\sin(-\theta) = -\sin(\theta) \quad \cos(-\theta) = \cos(\theta)$$

- Reflection around $\pi/4$

$$\sin\left(\frac{\pi}{2} - \theta\right) = \cos \theta \quad \cos\left(\frac{\pi}{2} - \theta\right) = \sin \theta$$

- Reflection around $\pi/2$

$$\sin(\pi - \theta) = \sin \theta \quad \cos(\pi - \theta) = -\cos \theta$$

- Rotation by π

$$\sin(\theta + \pi) = -\sin \theta \quad \cos(\theta + \pi) = -\cos \theta$$

- Pythagoras

$$\sin^2 \theta + \cos^2 \theta = 1$$

This page titled [A.7: Trigonometry — Simple Identities](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.8: Trigonometry — Add and Subtract Angles

- Sine

$$\sin(\alpha \pm \beta) = \sin(\alpha) \cos(\beta) \pm \cos(\alpha) \sin(\beta)$$

- Cosine

$$\cos(\alpha \pm \beta) = \cos(\alpha) \cos(\beta) \mp \sin(\alpha) \sin(\beta)$$

This page titled [A.8: Trigonometry — Add and Subtract Angles](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.9: Inverse Trigonometric Functions

Some of you may not have studied inverse trigonometric functions in highschool, however we still expect you to know them by the end of the course.



$\arcsin x$

Domain: $-1 \leq x \leq 1$

Range: $-\frac{\pi}{2} \leq \arcsin x \leq \frac{\pi}{2}$



$\arccos x$

Domain: $-1 \leq x \leq 1$

Range: $0 \leq \arccos x \leq \pi$



$\arctan x$

Domain: all real numbers

Range: $-\frac{\pi}{2} < \arctan x < \frac{\pi}{2}$

Since these functions are inverses of each other we have

$$\begin{aligned} \arcsin(\sin \theta) &= \theta & -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2} \\ \arccos(\cos \theta) &= \theta & 0 \leq \theta \leq \pi \\ \arctan(\tan \theta) &= \theta & -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2} \end{aligned}$$

and also

$$\begin{aligned} \sin(\arcsin x) &= x & -1 \leq x \leq 1 \\ \cos(\arccos x) &= x & -1 \leq x \leq 1 \\ \tan(\arctan x) &= x & \text{any real } x \end{aligned}$$



$\operatorname{arccsc} x$

Domain: $|x| \geq 1$

Range: $-\frac{\pi}{2} \leq \operatorname{arccsc} x \leq \frac{\pi}{2}$

$\operatorname{arccsc} x \neq 0$



$\operatorname{arcsec} x$

Domain: $|x| \geq 1$

Range: $0 \leq \operatorname{arcsec} x \leq \pi$

$\operatorname{arcsec} x \neq \frac{\pi}{2}$



$\operatorname{arccot} x$

Domain: all real numbers

Range: $0 < \operatorname{arccot} x < \pi$

Again

$$\begin{aligned} \operatorname{arccsc}(\csc \theta) &= \theta & -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}, \theta \neq 0 \\ \operatorname{arcsec}(\sec \theta) &= \theta & 0 \leq \theta \leq \pi, \theta \neq \frac{\pi}{2} \\ \operatorname{arccot}(\cot \theta) &= \theta & 0 < \theta < \pi \end{aligned}$$

and

$$\begin{aligned} \csc(\operatorname{arccsc} x) &= x & |x| \geq 1 \\ \sec(\operatorname{arcsec} x) &= x & |x| \geq 1 \\ \cot(\operatorname{arccot} x) &= x & \text{any real } x \end{aligned}$$

This page titled [A.9: Inverse Trigonometric Functions](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.10: Areas



- Area of a rectangle

$$A = bh$$

- Area of a triangle

$$A = \frac{1}{2}bh = \frac{1}{2}ab \sin \theta$$

- Area of a circle

$$A = \pi r^2$$

- Area of an ellipse

$$A = \pi ab$$

This page titled [A.10: Areas](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.11: Volumes

- Volume of a rectangular prism

$$V = lwh$$

- Volume of a cylinder

$$V = \pi r^2 h$$

- Volume of a cone

$$V = \frac{1}{3} \pi r^2 h$$

- Volume of a sphere

$$V = \frac{4}{3} \pi r^3$$

This page titled [A.11: Volumes](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.12: Powers

In the following, x and y are arbitrary real numbers, and q is an arbitrary constant that is strictly bigger than zero.

- $q^0 = 1$
- $q^{x+y} = q^x q^y$, $q^{x-y} = \frac{q^x}{q^y}$
- $q^{-x} = \frac{1}{q^x}$
- $(q^x)^y = q^{xy}$
- $\lim_{x \rightarrow \infty} q^x = \infty$, $\lim_{x \rightarrow -\infty} q^x = 0$ if $q > 1$
- $\lim_{x \rightarrow \infty} q^x = 0$, $\lim_{x \rightarrow -\infty} q^x = \infty$ if $0 < q < 1$
- The graph of 2^x is given below. The graph of q^x , for any $q > 1$, is similar.



This page titled [A.12: Powers](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.13: Logarithms

In the following, x and y are arbitrary real numbers that are strictly bigger than 0, and p and q are arbitrary constants that are strictly bigger than one.

- $q^{\log_q x} = x, \quad \log_q (q^x) = x$
- $\log_q x = \frac{\log_p x}{\log_p q}$
- $\log_q 1 = 0, \quad \log_q q = 1$
- $\log_q (xy) = \log_q x + \log_q y$
- $\log_q \left(\frac{x}{y}\right) = \log_q x - \log_q y$
- $\log_q \left(\frac{1}{y}\right) = -\log_q y,$
- $\log_q (x^y) = y \log_q x$
- $\lim_{x \rightarrow \infty} \log_q x = \infty, \quad \lim_{x \rightarrow 0} \log_q x = -\infty$
- The graph of $\log_{10} x$ is given below. The graph of $\log_q x$, for any $q > 1$, is similar.



This page titled [A.13: Logarithms](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.14: Highschool Material You Should be Able to Derive

- Graphs of $\csc \theta$, $\sec \theta$ and $\cot \theta$:



$\csc \theta$



$\sec \theta$



$\cot \theta$

- More Pythagoras

$$\begin{array}{l} \sin^2 \theta + \cos^2 \theta = 1 \quad \xrightarrow{\text{divide by } \cos^2 \theta} \quad \tan^2 \theta + 1 = \sec^2 \theta \\ \sin^2 \theta + \cos^2 \theta = 1 \quad \xrightarrow{\text{divide by } \sin^2 \theta} \quad 1 + \cot^2 \theta = \csc^2 \theta \end{array}$$

- Sine — double angle (set $\beta = \alpha$ in sine angle addition formula)

$$\sin(2\alpha) = 2 \sin(\alpha) \cos(\alpha)$$

- Cosine — double angle (set $\beta = \alpha$ in cosine angle addition formula)

$$\begin{aligned} \cos(2\alpha) &= \cos^2(\alpha) - \sin^2(\alpha) \\ &= 2 \cos^2(\alpha) - 1 && \text{(use } \sin^2(\alpha) = 1 - \cos^2(\alpha)) \\ &= 1 - 2 \sin^2(\alpha) && \text{(use } \cos^2(\alpha) = 1 - \sin^2(\alpha)) \end{aligned}$$

- Composition of trigonometric and inverse trigonometric functions:

$$\cos(\arcsin x) = \sqrt{1 - x^2} \quad \sec(\arctan x) = \sqrt{1 + x^2}$$

and similar expressions.

This page titled [A.14: Highschool Material You Should be Able to Derive](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.15: Cartesian Coordinates

Each point in two dimensions may be labeled by two coordinates (x, y) which specify the position of the point in some units with respect to some axes as in the figure below.



The set of all points in two dimensions is denoted \mathbb{R}^2 . Observe that

- the distance from the point (x, y) to the x -axis is $|y|$
- the distance from the point (x, y) to the y -axis is $|x|$
- the distance from the point (x, y) to the origin $(0, 0)$ is $\sqrt{x^2 + y^2}$

Similarly, each point in three dimensions may be labeled by three coordinates (x, y, z) , as in the two figures below.



The set of all points in three dimensions is denoted \mathbb{R}^3 . The plane that contains, for example, the x - and y -axes is called the xy -plane.

- The xy -plane is the set of all points (x, y, z) that obey $z = 0$.
- The xz -plane is the set of all points (x, y, z) that obey $y = 0$.
- The yz -plane is the set of all points (x, y, z) that obey $x = 0$.

More generally,

- The set of all points (x, y, z) that obey $z = c$ is a plane that is parallel to the xy -plane and is a distance $|c|$ from it. If $c > 0$, the plane $z = c$ is above the xy -plane. If $c < 0$, the plane $z = c$ is below the xy -plane. We say that the plane $z = c$ is a signed distance c from the xy -plane.
- The set of all points (x, y, z) that obey $y = b$ is a plane that is parallel to the xz -plane and is a signed distance b from it.
- The set of all points (x, y, z) that obey $x = a$ is a plane that is parallel to the yz -plane and is a signed distance a from it.



Observe that

- the distance from the point (x, y, z) to the xy -plane is $|z|$
- the distance from the point (x, y, z) to the xz -plane is $|y|$
- the distance from the point (x, y, z) to the yz -plane is $|x|$
- the distance from the point (x, y, z) to the origin $(0, 0, 0)$ is $\sqrt{x^2 + y^2 + z^2}$

The distance from the point (x, y, z) to the point (x', y', z') is

$$\sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}$$

so that the equation of the sphere centered on $(1, 2, 3)$ with radius 4, that is, the set of all points (x, y, z) whose distance from $(1, 2, 3)$ is 4, is

$$(x - 1)^2 + (y - 2)^2 + (z - 3)^2 = 16$$

This page titled [A.15: Cartesian Coordinates](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

A.16: Roots of Polynomials

Being able to factor polynomials is a very important part of many of the computations in this course. Related to this is the process of finding roots (or zeros) of polynomials. That is, given a polynomial $P(x)$, find all numbers r so that $P(r) = 0$.

In the case of a quadratic $P(x) = ax^2 + bx + c$, we can use the formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

The corresponding formulas for cubics and quartics ¹ are extremely cumbersome, and no such formula exists for polynomials of degree 5 and higher ².

Despite this there are many tricks ³ for finding roots of polynomials that work well in some situations but not all. Here we describe approaches that will help you find integer and rational roots of polynomials that will work well on exams, quizzes and homework assignments.

Consider the quadratic equation $x^2 - 5x + 6 = 0$. We could ⁴ solve this using the quadratic formula

$$x = \frac{5 \pm \sqrt{25 - 4 \times 1 \times 6}}{2} = \frac{5 \pm 1}{2} = 2, 3.$$

Hence $x^2 - 5x + 6$ has roots $x = 2, 3$ and so it factors as $(x - 3)(x - 2)$. Notice ⁵ that the numbers 2 and 3 divide the constant term of the polynomial, 6. This happens in general and forms the basis of our first trick.

Lemma A.16.1 A very useful trick

If r or $-r$ is an integer root of a polynomial $P(x) = a_n x^n + \dots + a_1 x + a_0$ with integer coefficients, then r is a factor of the constant term a_0 .

Proof

If r is a root of the polynomial we know that $P(r) = 0$. Hence

$$a_n \cdot r^n + \dots + a_1 \cdot r + a_0 = 0$$

If we isolate a_0 in this expression we get

$$a_0 = -[a_n r^n + \dots + a_1 r]$$

We can see that r divides every term on the right-hand side. This means that the right-hand side is an integer times r . Thus the left-hand side, being a_0 , is an integer times r , as required. The argument for when $-r$ is a root is almost identical.

Let us put this observation to work.

Example A.16.2 Integer roots of $x^3 - x^2 + 2$

Find the integer roots of $P(x) = x^3 - x^2 + 2$.

Solution:

- The constant term in this polynomial is 2.
- The only divisors of 2 are 1, 2. So the only candidates for integer roots are $\pm 1, \pm 2$.
- Trying each in turn

$$\begin{array}{ll} P(1) = 2 & P(-1) = 0 \\ P(2) = 6 & P(-2) = -10 \end{array}$$

- Thus the only integer root is -1 .

✓ Example A.16.3 Integer roots of $3x^3 + 8x^2 - 5x - 6$

Find the integer roots of $P(x) = 3x^3 + 8x^2 - 5x - 6$.

Solution:

- The constant term is -6 .
- The divisors of 6 are 1, 2, 3, 6. So the only candidates for integer roots are $\pm 1, \pm 2, \pm 3, \pm 6$.
- We try each in turn (it is tedious but not difficult):

$$\begin{array}{ll} P(1) = 0 & P(-1) = 4 \\ P(2) = 40 & P(-2) = 12 \\ P(3) = 132 & P(-3) = 0 \\ P(6) = 900 & P(-6) = -336 \end{array}$$

- Thus the only integer roots are 1 and -3 .

We can generalise this approach in order to find rational roots. Consider the polynomial $6x^2 - x - 2$. We can find its zeros using the quadratic formula:

$$x = \frac{1 \pm \sqrt{1+48}}{12} = \frac{1 \pm 7}{12} = -\frac{1}{2}, \frac{2}{3}.$$

Notice now that the numerators, 1 and 2, both divide the constant term of the polynomial (being 2). Similarly, the denominators, 2 and 3, both divide the coefficient of the highest power of x (being 6). This is quite general.

🔗 Lemma A.16.4 Another nice trick

If b/d or $-b/d$ is a rational root in lowest terms (i.e. b and d are integers with no common factors) of a polynomial $Q(x) = a_n x^n + \dots + a_1 x + a_0$ with integer coefficients, then the numerator b is a factor of the constant term a_0 and the denominator d is a factor of a_n .

Proof

Since $\frac{b}{d}$ is a root of $P(x)$ we know that

$$a_n (b/d)^n + \dots + a_1 (b/d) + a_0 = 0$$

Multiply this equation through by d^n to get

$$a_n b^n + \dots + a_1 b d^{n-1} + a_0 d^n = 0$$

Move terms around to isolate $a_0 d^n$:

$$a_0 d^n = -[a_n b^n + \dots + a_1 b d^{n-1}]$$

Now every term on the right-hand side is some integer times b . Thus the left-hand side must also be an integer times b . We know that d does not contain any factors of b , hence a_0 must be some integer times b (as required).

Similarly we can isolate the term $a_n b^n$:

$$a_n b^n = -[a_{n-1} b^{n-1} d + \dots + a_1 b d^{n-1} + a_0 d^n]$$

Now every term on the right-hand side is some integer times d . Thus the left-hand side must also be an integer times d . We know that b does not contain any factors of d , hence a_n must be some integer times d (as required).

The argument when $-\frac{b}{d}$ is a root is nearly identical.

We should put this to work:

✓ Example A.16.5 Rational roots of $2x^2 - x - 3$

$$P(x) = 2x^2 - x - 3.$$

Solution:

- The constant term in this polynomial is $3 = 1 \times 3$ and the coefficient of the highest power of x is $2 = 1 \times 2$.
- Thus the only candidates for integer roots are $\pm 1, \pm 3$.
- By our newest trick, the only candidates for fractional roots are $\pm \frac{1}{2}, \pm \frac{3}{2}$.
- We try each in turn ⁶

$$\begin{array}{ll} P(1) = -2 & P(-1) = 0 \\ P(3) = 12 & P(-3) = 18 \\ P\left(\frac{1}{2}\right) = -3 & P\left(-\frac{1}{2}\right) = -2 \\ P\left(\frac{3}{2}\right) = 0 & P\left(-\frac{3}{2}\right) = 3 \end{array}$$

so the roots are -1 and $\frac{3}{2}$.

The tricks above help us to find integer and rational roots of polynomials. With a little extra work we can extend those methods to help us factor polynomials. Say we have a polynomial $P(x)$ of degree p and have established that r is one of its roots. That is, we know $P(r) = 0$. Then we can factor $(x - r)$ out from $P(x)$ — it is always possible to find a polynomial $Q(x)$ of degree $p - 1$ so that

$$P(x) = (x - r)Q(x)$$

In sufficiently simple cases, you can probably do this factoring by inspection. For example, $P(x) = x^2 - 4$ has $r = 2$ as a root because $P(2) = 2^2 - 4 = 0$. In this case, $P(x) = (x - 2)(x + 2)$ so that $Q(x) = (x + 2)$. As another example, $P(x) = x^2 - 2x - 3$ has $r = -1$ as a root because $P(-1) = (-1)^2 - 2(-1) - 3 = 1 + 2 - 3 = 0$. In this case, $P(x) = (x + 1)(x - 3)$ so that $Q(x) = (x - 3)$.

For higher degree polynomials we need to use something more systematic — long division.

🔗 Lemma A.16.6 Long Division

Once you have found a root r of a polynomial, even if you cannot factor $(x - r)$ out of the polynomial by inspection, you can find $Q(x)$ by dividing $P(x)$ by $x - r$, using the long division algorithm you learned ⁷This is a standard part of most highschool mathematics curricula, but perhaps not all. You should revise this carefully. in school, but with 10 replaced by x .

✓ Example A.16.7 Roots of $x^3 - x^2 + 2$

Factor $P(x) = x^3 - x^2 + 2$.

Solution:

- We can go hunting for integer roots of the polynomial by looking at the divisors of the constant term. This tells us to try $x = \pm 1, \pm 2$.
- A quick computation shows that $P(-1) = 0$ while $P(1), P(-2), P(2) \neq 0$. Hence $x = -1$ is a root of the polynomial and so $x + 1$ must be a factor.
- So we divide $\frac{x^3 - x^2 + 2}{x + 1}$. The first term, x^2 , in the quotient is chosen so that when you multiply it by the denominator, $x^2(x + 1) = x^3 + x^2$, the leading term, x^3 , matches the leading term in the numerator, $x^3 - x^2 + 2$, exactly.

$$x + 1 \overline{) \begin{array}{r} x^3 - x^2 + 2 \\ x^3 + x^2 \\ \hline \end{array} \quad \leftarrow x^2(x + 1)$$

- When you subtract $x^2(x + 1) = x^3 + x^2$ from the numerator $x^3 - x^2 + 2$ you get the remainder $-2x^2 + 2$. Just like in public school, the 2 is not normally “brought down” until it is actually needed.

- The next term, $-2x$, in the quotient is chosen so that when you multiply it by the denominator, $-2x(x+1) = -2x^2 - 2x$, the leading term $-2x^2$ matches the leading term in the remainder exactly.

$$\begin{array}{r}
 x+1 \overline{) \begin{array}{r} x^3 - x^2 + 2 \\ x^3 + x^2 \\ \hline -2x^2 - 2x \end{array} } \leftarrow x^2(x+1) \\
 \phantom{x+1 \overline{) }} \leftarrow -2x(x+1)
 \end{array}$$

And so on.

- Note that we finally end up with a remainder 0. A nonzero remainder would have signalled a computational error, since we know that the denominator $x - (-1)$ must divide the numerator $x^3 - x^2 + 2$ exactly.
- We conclude that

$$(x+1)(x^2 - 2x + 2) = x^3 - x^2 + 2$$

To check this, just multiply out the left hand side explicitly.

- Applying the high school quadratic root formula $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ to $x^2 - 2x + 2$ tells us that it has no real roots and that we cannot factor it further⁸.

We finish by describing an alternative to long division. The approach is roughly equivalent, but is perhaps more straightforward at the expense of requiring more algebra.

✓ Example A.16.8 Roots of $x^3 - x^2 + 2$ again

Factor $P(x) = x^3 - x^2 + 2$, again.

Solution: Let us do this again but avoid long division.

- From the previous example, we know that $\frac{x^3 - x^2 + 2}{x+1}$ must be a polynomial (since -1 is a root of the numerator) of degree 2. So write

$$\frac{x^3 - x^2 + 2}{x+1} = ax^2 + bx + c$$

for some, as yet unknown, coefficients a , b and c .

- Cross multiplying and simplifying gives us

$$\begin{aligned}
 x^3 - x^2 + 2 &= (ax^2 + bx + c)(x+1) \\
 &= ax^3 + (a+b)x^2 + (b+c)x + c
 \end{aligned}$$

- Now matching coefficients of the various powers of x on the left and right hand sides

$$\begin{array}{ll}
 \text{coefficient of } x^3: & a = 1 \\
 \text{coefficient of } x^2: & a + b = -1 \\
 \text{coefficient of } x^1: & b + c = 0 \\
 \text{coefficient of } x^0: & c = 2
 \end{array}$$

- This gives us a system of equations that we can solve quite directly. Indeed it tells us immediately that that $a = 1$ and $c = 2$. Subbing $a = 1$ into $a + b = -1$ tells us that $1 + b = -1$ and hence $b = -2$.

- Thus

$$x^3 - x^2 + 2 = (x + 1)(x^2 - 2x + 2).$$

This page titled [A.16: Roots of Polynomials](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

B: Complex Numbers and Exponentials

This page titled [B: Complex Numbers and Exponentials](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

B.1: Definition and Basic Operations

We'll start with the definition of a complex number and its addition and multiplication rules. You may find the multiplication rule quite mysterious. Don't worry. We'll soon get lots of practice using it and seeing how useful it is.

Definition B.1.1

1. The complex plane is simply the xy -plane equipped with an addition operation and a multiplication operation. A complex number is nothing more than a point in that xy -plane. It is conventional to use the notation $x + iy$ to stand for the complex number (x, y) . In other words, it is conventional to write x in place of $(x, 0)$ and i in place of $(0, 1)$.
2. The first component, x , of the complex number $x + iy$ is called its real part and the second component, y , is called its imaginary part, even though there is nothing imaginary² about it.
3. The sum of the complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ is defined by

$$z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2)$$

That is, you just separately add the real parts and the imaginary parts.

4. The product of the complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ is defined by

$$z_1 z_2 = x_1 x_2 - y_1 y_2 + i(x_1 y_2 + x_2 y_1)$$

Do not memorise this multiplication rule. We'll see a simple, effective memory aid for it shortly. The heart of that memory aid is the observation that the complex number i has the special property that

$$i^2 = (0 + 1i)(0 + 1i) = (0 \times 0 - 1 \times 1) + i(0 \times 1 + 1 \times 0) = -1$$

Addition and multiplication of complex numbers obey the familiar addition rules

$$\begin{aligned} z_1 + z_2 &= z_2 + z_1 \\ z_1 + (z_2 + z_3) &= (z_1 + z_2) + z_3 \\ 0 + z_1 &= z_1 \end{aligned}$$

and multiplication rules

$$\begin{aligned} z_1 z_2 &= z_2 z_1 \\ z_1(z_2 z_3) &= (z_1 z_2)z_3 \\ 1z_1 &= z_1 \end{aligned}$$

and distributive laws

$$\begin{aligned} z_1(z_2 + z_3) &= z_1 z_2 + z_1 z_3 \\ (z_1 + z_2)z_3 &= z_1 z_3 + z_2 z_3 \end{aligned}$$

To remember how to multiply complex numbers, you just have to supplement the familiar rules of the real number system with $i^2 = -1$. The previous sentence is the memory aid referred to in Definition B.1.1(d).

Example B.1.2

If $z = 1 + 2i$ and $w = 3 + 4i$, then

$$\begin{aligned} z + w &= (1 + 2i) + (3 + 4i) = 4 + 6i \\ zw &= (1 + 2i)(3 + 4i) = 3 + 4i + 6i + 8i^2 = 3 + 4i + 6i - 8 \\ &= -5 + 10i \end{aligned}$$

Definition B.1.3

1. The negative of any complex number $z = x + iy$ is defined by

$$-z = -x + (-y)i$$

and obviously obeys $z + (-z) = 0$.

2. The reciprocal³, z^{-1} or $\frac{1}{z}$, of any complex number $z = x + iy$, other than 0, is defined by

$$\frac{1}{z}z = 1$$

We shall see below that it is given by the formula

$$\frac{1}{z} = \frac{x}{x^2 + y^2} + \frac{-y}{x^2 + y^2}i$$

✓ Example B.1.4

It is possible to derive the formula for $\frac{1}{z}$ by observing that

$$(a + ib)(x + iy) = [ax - by] + i[ay + bx]$$

equals $1 = 1 + i0$ if and only if

$$\begin{aligned} ax - by &= 1 \\ ay + bx &= 0 \end{aligned}$$

and solving these equations for a and b . We will see a much shorter derivation in Remark B.1.6 below. For now, we'll content ourselves with just verifying that $\frac{x}{x^2 + y^2} + \frac{-y}{x^2 + y^2}i$ is the inverse of $x + iy$ by multiplying out

$$\begin{aligned} &\left(\frac{x}{x^2 + y^2} - \frac{y}{x^2 + y^2}i\right)(x + iy) \\ &= \frac{x^2}{x^2 + y^2} - \frac{xy}{x^2 + y^2}i + \frac{xy}{x^2 + y^2}i - \frac{y^2}{x^2 + y^2}i^2 \\ &= \frac{x^2 - i^2y^2}{x^2 + y^2} = \frac{x^2 + y^2}{x^2 + y^2} = 1 \end{aligned}$$

Definition B.1.5

1. The complex conjugate of $z = x + iy$ is denoted \bar{z} and is defined to be

$$\bar{z} = \overline{x + iy} = x - iy$$

That is, to take the complex conjugate, one replaces every i by $-i$ and vice versa.

2. The distance from $z = x + iy$ (recall that this is the point (x, y) in the xy -plane) to 0 is denoted $|z|$ and is called the absolute value, or modulus, of z . It is given by

$$|z| = \sqrt{x^2 + y^2}$$

Note that

$$z\bar{z} = (x + iy)(x - iy) = x^2 - ixy + ixy + y^2 = x^2 + y^2$$

is always a nonnegative real number and that

$$|z| = \sqrt{z\bar{z}}$$

Remark B.1.6

Let $z = x + iy$ with x and y real. Since $|z|^2 = z\bar{z}$, we have

$$\frac{1}{z} = \frac{1}{z\bar{z}} = \frac{\bar{z}}{|z|^2} = \frac{x - iy}{x^2 + y^2} = \frac{x}{x^2 + y^2} + \frac{-y}{x^2 + y^2}i$$

which is the formula for $\frac{1}{z}$ given in Definition B.1.3(b).

✓ Example B.1.7

It is easy to divide a complex number by a real number. For example

$$\frac{11 + 2i}{25} = \frac{11}{25} + \frac{2}{25}i$$

In general, the complex conjugate provides us with a trick for rewriting any ratio of complex numbers as a ratio with a real denominator. For example, suppose that we want to find $\frac{1+2i}{3+4i}$. The trick is to multiply by $1 = \frac{3-4i}{3-4i}$. The number $3 - 4i$ is the complex conjugate of the denominator $3 + 4i$. Since $(3 + 4i)(3 - 4i) = 9 - 12i + 12i - 16i^2 = 9 + 16 = 25$

$$\frac{1 + 2i}{3 + 4i} = \frac{1 + 2i}{3 + 4i} \frac{3 - 4i}{3 - 4i} = \frac{(1 + 2i)(3 - 4i)}{25} = \frac{11 + 2i}{25} = \frac{11}{25} + \frac{2}{25}i$$

✎ Definition B.1.8

The notations $\Re z$ and $\Im z$ stand for the real and imaginary parts of the complex number z , respectively. If $z = x + iy$ (with x and y real) they are defined by

$$\Re z = x \quad \Im z = y$$

Note that both $\Re z$ and $\Im z$ are real numbers. Just subbing in $\bar{z} = x - iy$, you can verify that

$$\Re z = \frac{1}{2}(z + \bar{z}) \quad \Im z = \frac{1}{2i}(z - \bar{z})$$

✎ Lemma B.1.9

If $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$, then

$$|z_1 z_2| = |z_1| |z_2|$$

Proof

Since $z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1)$,

$$\begin{aligned} |z_1 z_2| &= \sqrt{(x_1 x_2 - y_1 y_2)^2 + (x_1 y_2 + x_2 y_1)^2} \\ &= \sqrt{x_1^2 x_2^2 - 2x_1 x_2 y_1 y_2 + y_1^2 y_2^2 + x_1^2 y_2^2 + 2x_1 y_2 x_2 y_1 + x_2^2 y_1^2} \\ &= \sqrt{x_1^2 x_2^2 + y_1^2 y_2^2 + x_1^2 y_2^2 + x_2^2 y_1^2} = \sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)} \\ &= |z_1| |z_2| \end{aligned}$$

This page titled [B.1: Definition and Basic Operations](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

B.2: The Complex Exponential

Definition and Basic Properties

There are two equivalent standard definitions of the exponential, e^z , of the complex number $z = x + iy$. For the more intuitive definition, one simply replaces the real number x in the Taylor series expansion¹ $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ with the complex number z , giving

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

We instead highlight the more computationally useful definition.

Definition B.2.1

For any complex number $z = x + iy$, with x and y real, the exponential e^z , is defined by

$$e^{x+iy} = e^x \cos y + i e^x \sin y$$

In particular², $e^{iy} = \cos y + i \sin y$.

We will not fully prove that the intuitive definition (EZ) and the computational Definition B.2.1 are equivalent. But we will do so in the special case that $z = iy$, with y real. Under (EZ),

$$e^{iy} = 1 + iy + \frac{(iy)^2}{2!} + \frac{(iy)^3}{3!} + \frac{(iy)^4}{4!} + \frac{(iy)^5}{5!} + \frac{(iy)^6}{6!} + \dots$$

The even terms in this expansion are

$$1 + \frac{(iy)^2}{2!} + \frac{(iy)^4}{4!} + \frac{(iy)^6}{6!} + \dots = 1 - \frac{y^2}{2!} + \frac{y^4}{4!} - \frac{y^6}{6!} + \dots = \cos y$$

and the odd terms in this expansion are

$$iy + \frac{(iy)^3}{3!} + \frac{(iy)^5}{5!} + \dots = i \left(y - \frac{y^3}{3!} + \frac{y^5}{5!} + \dots \right) = i \sin y$$

Adding the even and odd terms together gives us that, under (EZ), e^{iy} is indeed equal to $\cos y + i \sin y$.³ It is obvious that, in the special case that $z = x$ with x real, the definitions (EZ) and B.2.1 are equivalent. So to complete the proof of equivalence in the general case $z = x + iy$, it suffices to prove that $e^{x+iy} = e^x e^{iy}$ under both (EZ) and Definition B.2.1. For Definition B.2.1, this follows from Lemma B.2.2, below. As a consequence, we have

$$e^{i\pi} = -1$$

which gives an amazing linking between calculus (e), geometry (π), algebra (i) and the basic number -1 .

In the next lemma we verify that the complex exponential obeys a couple of familiar computational properties.

Lemma B.2.2

1. For any complex numbers z_1 and z_2 ,

$$e^{z_1+z_2} = e^{z_1} e^{z_2}$$

2. For any complex number c ,

$$\frac{d}{dt} e^{ct} = c e^{ct}$$

Proof

1. For any two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$, with x_1, y_1, x_2, y_2 real,

$$\begin{aligned}
 e^{z_1} e^{z_2} &= e^{x_1} (\cos y_1 + i \sin y_1) e^{x_2} (\cos y_2 + i \sin y_2) \\
 &= e^{x_1+x_2} (\cos y_1 + i \sin y_1) (\cos y_2 + i \sin y_2) \\
 &= e^{x_1+x_2} \{ (\cos y_1 \cos y_2 - \sin y_1 \sin y_2) + i (\cos y_1 \sin y_2 + \cos y_2 \sin y_1) \} \\
 &= e^{x_1+x_2} \{ \cos(y_1 + y_2) + i \sin(y_1 + y_2) \}
 \end{aligned}$$

by the trig identities of Appendix A.8. This says exactly that

$$e^{z_1} e^{z_2} = e^{(x_1+x_2)+i(y_1+y_2)} = e^{z_1+z_2}$$

and that the familiar multiplication formula also applies to complex exponentials.

2. For any real number t and any complex number $c = \alpha + i\beta$, with α, β real,

$$e^{ct} = e^{\alpha t + i\beta t} = e^{\alpha t} [\cos(\beta t) + i \sin(\beta t)]$$

so that the derivative with respect to t

$$\begin{aligned}
 \frac{d}{dt} e^{ct} &= \alpha e^{\alpha t} [\cos(\beta t) + i \sin(\beta t)] + e^{\alpha t} [-\beta \sin(\beta t) + i\beta \cos(\beta t)] \\
 &= (\alpha + i\beta) e^{\alpha t} [\cos(\beta t) + i \sin(\beta t)] \\
 &= c e^{ct}
 \end{aligned}$$

is also the familiar one.

Relationship with **sin** and **cos**

Equation B.2.3

When θ is a real number

$$\begin{aligned}
 e^{i\theta} &= \cos \theta + i \sin \theta \\
 e^{-i\theta} &= \cos \theta - i \sin \theta = \overline{e^{i\theta}}
 \end{aligned}$$

are complex numbers of modulus one.

Solving for $\cos \theta$ and $\sin \theta$ (by adding and subtracting the two equations) gives

Equation B.2.4

$$\begin{aligned}
 \cos \theta &= \frac{1}{2} (e^{i\theta} + e^{-i\theta}) = \Re e^{i\theta} \\
 \sin \theta &= \frac{1}{2i} (e^{i\theta} - e^{-i\theta}) = \Im e^{i\theta}
 \end{aligned}$$

Example B.2.5

These formulae make it easy derive trig identities. For example,

$$\begin{aligned}
 \cos \theta \cos \phi &= \frac{1}{4} (e^{i\theta} + e^{-i\theta}) (e^{i\phi} + e^{-i\phi}) \\
 &= \frac{1}{4} (e^{i(\theta+\phi)} + e^{i(\theta-\phi)} + e^{i(-\theta+\phi)} + e^{-i(\theta+\phi)}) \\
 &= \frac{1}{4} (e^{i(\theta+\phi)} + e^{-i(\theta+\phi)} + e^{i(\theta-\phi)} + e^{i(-\theta+\phi)}) \\
 &= \frac{1}{2} (\cos(\theta + \phi) + \cos(\theta - \phi))
 \end{aligned}$$

and, using $(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$,

$$\begin{aligned}\sin^3 \theta &= -\frac{1}{8i}(e^{i\theta} - e^{-i\theta})^3 \\ &= -\frac{1}{8i}(e^{i3\theta} - 3e^{i\theta} + 3e^{-i\theta} - e^{-i3\theta}) \\ &= \frac{3}{4} \frac{1}{2i}(e^{i\theta} - e^{-i\theta}) - \frac{1}{4} \frac{1}{2i}(e^{i3\theta} - e^{-i3\theta}) \\ &= \frac{3}{4} \sin \theta - \frac{1}{4} \sin(3\theta)\end{aligned}$$

and

$$\begin{aligned}\cos(2\theta) &= \Re(e^{2\theta i}) = \Re(e^{i\theta})^2 \\ &= \Re(\cos \theta + i \sin \theta)^2 \\ &= \Re(\cos^2 \theta + 2i \sin \theta \cos \theta - \sin^2 \theta) \\ &= \cos^2 \theta - \sin^2 \theta\end{aligned}$$

Polar Coordinates

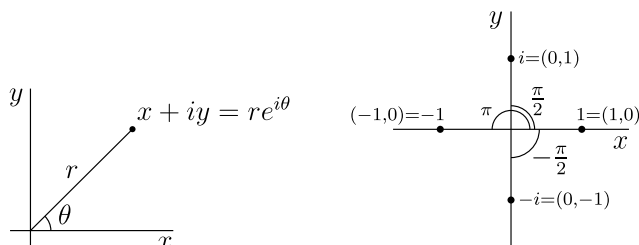
Let $z = x + iy$ be any complex number. Writing x and y in polar coordinates in the usual way, i.e. $x = r \cos(\theta)$, $y = r \sin(\theta)$, gives

$$x + iy = r \cos \theta + ir \sin \theta = re^{i\theta}$$

See the figure on the left below. In particular

$$\begin{aligned}1 &= e^{i0} = e^{2\pi i} = e^{2k\pi i} && \text{for } k = 0, \pm 1, \pm 2, \dots \\ -1 &= e^{i\pi} = e^{3\pi i} = e^{(1+2k)\pi i} && \text{for } k = 0, \pm 1, \pm 2, \dots \\ i &= e^{i\pi/2} = e^{\frac{5}{2}\pi i} = e^{(\frac{1}{2}+2k)\pi i} && \text{for } k = 0, \pm 1, \pm 2, \dots \\ -i &= e^{-i\pi/2} = e^{\frac{3}{2}\pi i} = e^{(-\frac{1}{2}+2k)\pi i} && \text{for } k = 0, \pm 1, \pm 2, \dots\end{aligned}$$

See the figure on the right below.



The polar coordinate $\theta = \arctan \frac{y}{x}$ associated with the complex number $z = x + iy$, i.e. the point (x, y) in the xy -plane, is also called the argument of z .

The polar coordinate representation makes it easy to find square roots, third roots and so on. Fix any positive integer n . The n^{th} roots of unity are, by definition, all solutions z of

$$z^n = 1$$

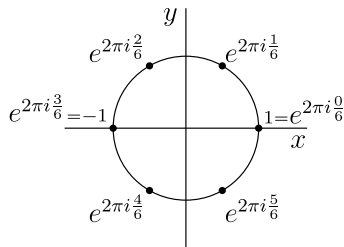
Writing $z = re^{i\theta}$

$$r^n e^{n\theta i} = 1e^{0i}$$

The polar coordinates (r, θ) and (r', θ') represent the same point in the xy -plane if and only if $r = r'$ and $\theta = \theta' + 2k\pi$ for some integer k . So $z^n = 1$ if and only if $r^n = 1$, i.e. $r = 1$, and $n\theta = 2k\pi$ for some integer k . The n^{th} roots of unity are all the complex numbers $e^{2\pi i \frac{k}{n}}$ with k integer. There are precisely n distinct n^{th} roots of unity because $e^{2\pi i \frac{k}{n}} = e^{2\pi i \frac{k'}{n}}$ if and only if $2\pi \frac{k}{n} - 2\pi \frac{k'}{n} = 2\pi \frac{k-k'}{n}$ is an integer multiple of 2π . That is, if and only if $k - k'$ is an integer multiple of n . The n distinct n^{th} roots of unity are

$$1, e^{2\pi i \frac{1}{n}}, e^{2\pi i \frac{2}{n}}, e^{2\pi i \frac{3}{n}}, \dots, e^{2\pi i \frac{n-1}{n}}$$

For example, the 6th roots of unity are depicted below.



Exploiting Complex Exponentials in Calculus Computations

You have learned how to evaluate integrals involving trigonometric functions by using integration by parts, various trigonometric identities and various substitutions. It is often much easier to just use (B.2.3) and (B.2.4). Part of the utility of complex numbers comes from how well they interact with calculus through the exponential function. Here are two examples.

✓ Example B.2.6

$$\begin{aligned} \int e^x \cos x \, dx &= \frac{1}{2} \int e^x [e^{ix} + e^{-ix}] \, dx = \frac{1}{2} \int [e^{(1+i)x} + e^{(1-i)x}] \, dx \\ &= \frac{1}{2} \left[\frac{1}{1+i} e^{(1+i)x} + \frac{1}{1-i} e^{(1-i)x} \right] + C \end{aligned}$$

This form of the indefinite integral looks a little weird because of the i 's. While it looks complex because of the i 's, it is actually purely real (and correct), because $\frac{1}{1-i} e^{(1-i)x}$ is the complex conjugate of $\frac{1}{1+i} e^{(1+i)x}$. We can convert the indefinite integral into a more familiar form just by subbing back in $e^{\pm ix} = \cos x \pm i \sin x$, $\frac{1}{1+i} = \frac{1-i}{(1+i)(1-i)} = \frac{1-i}{2}$ and $\frac{1}{1-i} = \frac{1+i}{2}$.

$$\begin{aligned} \int e^x \cos x \, dx &= \frac{1}{2} e^x \left[\frac{1}{1+i} e^{ix} + \frac{1}{1-i} e^{-ix} \right] + C \\ &= \frac{1}{2} e^x \left[\frac{1-i}{2} (\cos x + i \sin x) + \frac{1+i}{2} (\cos x - i \sin x) \right] + C \\ &= \frac{1}{2} e^x \cos x + \frac{1}{2} e^x \sin x + C \end{aligned}$$

You can quickly verify this by differentiating (or by comparing with Example 1.7.11).

✓ Example B.2.7

Evaluating the integral $\int \cos^n x \, dx$ using the methods of Section 1.8 can be a real pain. It is much easier if we convert to complex exponentials. Using $(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$,

$$\begin{aligned}
 \int \cos^4 x \, dx &= \frac{1}{2^4} \int [e^{ix} + e^{-ix}]^4 \, dx \\
 &= \frac{1}{2^4} \int [e^{4ix} + 4e^{2ix} + 6 + 4e^{-2ix} + e^{-4ix}] \, dx \\
 &= \frac{1}{2^4} \left[\frac{1}{4i} e^{4ix} + \frac{4}{2i} e^{2ix} + 6x + \frac{4}{-2i} e^{-2ix} + \frac{1}{-4i} e^{-4ix} \right] + C \\
 &= \frac{1}{2^4} \left[\frac{1}{2} \frac{1}{2i} (e^{4ix} - e^{-4ix}) + \frac{4}{2i} (e^{2ix} - e^{-2ix}) + 6x \right] + C \\
 &= \frac{1}{2^4} \left[\frac{1}{2} \sin 4x + 4 \sin 2x + 6x \right] + C \\
 &= \frac{1}{32} \sin 4x + \frac{1}{4} \sin 2x + \frac{3}{8} x + C
 \end{aligned}$$

Exploiting Complex Exponentials in Differential Equation Computations

Complex exponentials are also widely used to simplify the process of guessing solutions to ordinary differential equations. We'll start with (possibly a review of) some basic definitions and facts about differential equations.

Definition B.2.8

1. A **differential equation** is an equation for an unknown function that contains the derivatives of that unknown function. For example $y''(t) + y(t) = 0$ is a differential equation for the unknown function $y(t)$.
2. In the differential calculus text CLP-1, we treated only derivatives of functions of one variable. Such derivatives are called ordinary derivatives. A differential equation is called an **ordinary differential equation** (often shortened to “ODE”) if only ordinary derivatives appear. That is, if the unknown function has only a single independent variable.

In CLP-3 we will treat derivatives of functions of more than one variable. For example, let $f(x, y)$ be a function of two variables. If you treat y as a constant and take the derivative of the resulting function of the single variable x , the result is called the partial derivative of f with respect to x . A differential equation is called a **partial differential equation** (often shortened to “PDE”) if partial derivatives appear. That is, if the unknown function has more than one independent variable.

For example $y''(t) + y(t) = 0$ is an ODE while $\frac{\partial^2 u}{\partial t^2}(x, t) = c^2 \frac{\partial^2 u}{\partial x^2}(x, t)$ is a PDE.

3. The **order** of a differential equation is the order of the highest derivative that appears. For example $y''(t) + y(t) = 0$ is a second order ODE.
4. An ordinary differential equation that is of the form

$$a_0(t)y^{(n)}(t) + a_1(t)y^{(n-1)}(t) + \cdots + a_n(t)y(t) = F(t) \quad (\text{B.2.1})$$

with given coefficient functions $a_0(t), \dots, a_n(t)$ and $F(t)$ is said to be **linear**. Otherwise, the ODE is said to be **nonlinear**. For example, $y'(t)^2 + y(t) = 0$, $y'(t)y''(t) + y(t) = 0$ and $y'(t) = e^{y(t)}$ are all nonlinear.

5. The ODE (B.2.1) is said to have **constant coefficients** if the coefficients $a_0(t), a_1(t), \dots, a_n(t)$ are all constants. Otherwise, it is said to have **variable coefficients**. For example, the ODE $y''(t) + 7y(t) = \sin t$ is constant coefficient, while $y''(t) + ty(t) = \sin t$ is variable coefficient.
6. The ODE (B.2.1) is said to be **homogeneous** if $F(t)$ is identically zero. Otherwise, it is said to be **inhomogeneous** or **nonhomogeneous**. For example, the ODE $y''(t) + 7y(t) = 0$ is homogeneous, while $y''(t) + 7y(t) = \sin t$ is inhomogeneous. A homogeneous ODE always has the trivial solution $y(t) = 0$.
7. An **initial value problem** is a problem in which one is to find an unknown function $y(t)$ that satisfies both a given ODE and given initial conditions, like $y(t_0) = 1$, $y'(t_0) = 0$. Note that all of the conditions involve the function $y(t)$ (or its derivatives) evaluated at a single time $t = t_0$.
8. A **boundary value problem** is a problem in which one is to find an unknown function $y(t)$ that satisfies both a given ODE and given boundary conditions, like $y(t_0) = 0$, $y(t_1) = 0$. Note that the conditions involve the function $y(t)$ (or its derivatives) evaluated at two different times.

The following theorem gives the form of solutions to the linear⁴ ODE (B.2.1).

Theorem B.2.9

Assume that the coefficients $a_0(t), a_1(t), \dots, a_{n-1}(t), a_n(t)$ and $F(t)$ are continuous functions and that $a_0(t)$ is not zero.

1. The general solution to the linear ODE (B.2.1) is of the form

$$y(t) = y_p(t) + C_1 y_1(t) + C_2 y_2(t) + \dots + C_n y_n(t) \quad (\text{B.2.2})$$

where

- n is the order of (B.2.1)
- $y_p(t)$ is **any** solution to (B.2.1)
- C_1, C_2, \dots, C_n are arbitrary constants
- y_1, y_2, \dots, y_n are n independent solutions to the homogenous equation

$$a_0(t)y^{(n)}(t) + a_1(t)y^{(n-1)}(t) + \dots + a_{n-1}(t)y'(t) + a_n(t)y(t) = 0$$

associated to (B.2.1). “Independent” just means that no y_i can be written as a linear combination of the other y_j 's. For example, $y_1(t)$ cannot be expressed in the form $b_2 y_2(t) + \dots + b_n y_n(t)$.

In (B.2.2), y_p is called the “particular solution” and $C_1 y_1(t) + C_2 y_2(t) + \dots + C_n y_n(t)$ is called the “complementary solution”.

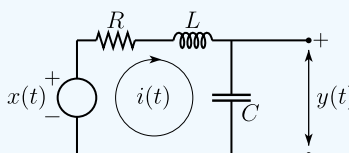
2. Given any constants b_0, \dots, b_{n-1} there is exactly one function $y(t)$ that obeys the ODE (B.2.1) and the initial conditions

$$y(0) = b_0 \quad y'(0) = b_1 \quad \dots \quad y^{(n-1)}(0) = b_{n-1}$$

In the following example we'll derive one widely used linear constant coefficient ODE.

✓ Example B.2.10 RLC circuit

As an example of how ODE's arise, we consider the RLC circuit, which is the electrical circuit consisting of a resistor of resistance R , a coil (or solenoid) of inductance L , a capacitor of capacitance C and a voltage source arranged in series, as shown below. Here R, L and C are all nonnegative constants.



We're going to think of the voltage $x(t)$ as an input signal, and the voltage $y(t)$ as an output signal. The goal is to determine the output signal produced by a given input signal. If $i(t)$ is the current flowing at time t in the loop as shown and $q(t)$ is the charge on the capacitor, then the voltages across R, L and C , respectively, at time t are $Ri(t), L \frac{di}{dt}(t)$ and $y(t) = \frac{q(t)}{C}$. By the Kirchhoff's law⁵ that says that the voltage between any two points has to be independent of the path used to travel between the two points, these three voltages must add up to $x(t)$ so that

$$Ri(t) + L \frac{di}{dt}(t) + \frac{q(t)}{C} = x(t) \quad (\text{B.2.3})$$

Assuming that R, L, C and $x(t)$ are known, this is still one differential equation in two unknowns, the current $i(t)$ and the charge $q(t)$. Fortunately, there is a relationship between the two. Because the current entering the capacitor is the rate of change of the charge on the capacitor

$$i(t) = \frac{dq}{dt}(t) = Cy'(t) \quad (\text{B.2.4})$$

This just says that the capacitor cannot create or destroy charge on its own; all charging of the capacitor must come from the current. Substituting (B.2.4) into (B.2.3) gives

$$LCy''(t) + RCy'(t) + y(t) = x(t)$$

which is a second order linear constant coefficient ODE. As a concrete example, we'll take an ac voltage source and choose the origin of time so that $x(0) = 0$, $x(t) = E_0 \sin(\omega t)$. Then the differential equation becomes

$$LCy''(t) + RCy'(t) + y(t) = E_0 \sin(\omega t) \quad (\text{B.2.5})$$

Finally, here are two examples in which we use complex exponentials to solve an ODE.

✓ Example B.2.11

By Theorem B.2.9(a), the general solution to the ordinary differential equation

$$y''(t) + 4y'(t) + 5y(t) = 0$$

is of the form $C_1 u_1(t) + C_2 u_2(t)$ with $u_1(t)$ and $u_2(t)$ being two (independent) solutions to (ODE) and with C_1 and C_2 being arbitrary constants. The easiest way to find $u_1(t)$ and $u_2(t)$ is to guess them. And the easiest way to guess them is to try ⁶The reason that $y(t) = e^{rt}$ is a good guess is that, with this guess, all of $y(t)$, $y'(t)$ and $y''(t)$ are constants times e^{rt} . So the left hand side of the differential equation is also a constant, that depends on r , times e^{rt} . So we just have to choose r so that the constant is zero. $y(t) = e^{rt}$, with r being a constant to be determined. Substituting $y(t) = e^{rt}$ into (ODE) gives

$$r^2 e^{rt} + 4r e^{rt} + 5e^{rt} = 0 \iff (r^2 + 4r + 5)e^{rt} = 0 \iff r^2 + 4r + 5 = 0$$

This quadratic equation for r can be solved either by using the high school formula or by completing the square.

$$\begin{aligned} r^2 + 4r + 5 = 0 &\iff (r+2)^2 + 1 = 0 \\ &\iff (r+2)^2 = -1 \iff r+2 = \pm i \\ &\iff r = -2 \pm i \end{aligned}$$

So the general solution to (ODE) is

$$y(t) = C_1 e^{(-2+i)t} + C_2 e^{(-2-i)t}$$

This is one way to write the general solution, but there are many others. In particular there are quite a few people in the world who are (foolishly) afraid ⁷Embracing the complexity leads to simplicity. of complex exponentials. We can hide them by using (B.2.3) and (B.2.4).

$$\begin{aligned} y(t) &= C_1 e^{(-2+i)t} + C_2 e^{(-2-i)t} = C_1 e^{-2t} e^{it} + C_2 e^{-2t} e^{-it} \\ &= C_1 e^{-2t} (\cos t + i \sin t) + C_2 e^{-2t} (\cos t - i \sin t) \\ &= (C_1 + C_2) e^{-2t} \cos t + (iC_1 - iC_2) e^{-2t} \sin t \\ &= D_1 e^{-2t} \cos t + D_2 e^{-2t} \sin t \end{aligned}$$

with $D_1 = C_1 + C_2$ and $D_2 = iC_1 - iC_2$ being two other arbitrary constants. Don't make the mistake of thinking that D_2 must be complex because i appears in the formula $D_2 = iC_1 - iC_2$ relating D_2 and C_1, C_2 . No one said that C_1 and C_2 are real numbers. In fact, in typical applications, the arbitrary constants are determined by initial conditions and often D_1 and D_2 turn out to be real and C_1 and C_2 turn out to be complex. For example, the initial conditions $y(0) = 0$, $y'(0) = 2$ force

$$\begin{aligned} 0 &= y(0) = C_1 + C_2 \\ 2 &= y'(0) = (-2+i)C_1 + (-2-i)C_2 \end{aligned}$$

The first equation gives $C_2 = -C_1$ and then the second equation gives

$$\begin{aligned} (-2+i)C_1 - (-2-i)C_1 &= 2 \iff 2iC_1 = 2 \iff iC_1 = 1 \\ &\iff C_1 = -i, C_2 = i \end{aligned}$$

and

$$D_1 = C_1 + C_2 = 0 \quad D_2 = iC_1 - iC_2 = 2$$

✓ Example B.2.12

We shall now guess one solution (i.e. a particular solution) to the differential equation

$$y''(t) + 2y'(t) + 3y(t) = \cos t$$

Equations like this arise, for example, in the study of the RLC circuit. We shall simplify the computation by exploiting that $\cos t = \Re e^{it}$. First, we shall guess a function $Y(t)$ obeying

$$Y'' + 2Y' + 3Y = e^{it}$$

Then, taking complex conjugates gives

$$\bar{Y}'' + 2\bar{Y}' + 3\bar{Y} = e^{-it}$$

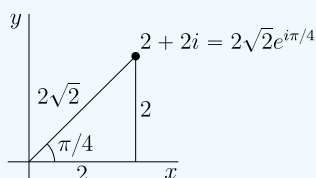
which we shall call $(\overline{\text{ODE2}})$. Then, adding $\frac{1}{2}(\text{ODE2})$ and $\frac{1}{2}(\overline{\text{ODE2}})$ together will give

$$(\Re Y)'' + 2(\Re Y)' + 3(\Re Y) = \Re e^{it} = \cos t$$

which shows that $\Re Y(t)$ is a solution to (ODE1). Let's try $Y(t) = Ae^{it}$, with A a constant to be determined. This is a solution of (ODE2) if and only if

$$\begin{aligned} \frac{d^2}{dt^2}(Ae^{it}) + 2\frac{d}{dt}(Ae^{it}) + 3Ae^{it} &= e^{it} \\ \iff (i^2 + 2i + 3)Ae^{it} &= e^{it} \\ \iff A &= \frac{1}{2 + 2i} \end{aligned}$$

So $\frac{e^{it}}{2+2i}$ is a solution to (ODE2) and $\Re \frac{e^{it}}{2+2i}$ is a solution to (ODE1). To simplify this, write $2 + 2i$ in polar coordinates. From the sketch below we have $2 + 2i = 2\sqrt{2}e^{i\pi/4}$.



So

$$\begin{aligned} \frac{e^{it}}{2 + 2i} &= \frac{e^{it}}{2\sqrt{2}e^{i\pi/4}} = \frac{1}{2\sqrt{2}}e^{i(t-\pi/4)} \\ \implies \Re \frac{e^{it}}{2 + 2i} &= \frac{1}{2\sqrt{2}}\Re e^{i(t-\pi/4)} = \frac{1}{2\sqrt{2}}\cos\left(t - \frac{\pi}{4}\right) \end{aligned}$$

C: More About Numerical Integration

This page titled [C: More About Numerical Integration](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

C.1: Richardson Extrapolation

There are many approximation procedures in which one first picks a step size h and then generates an approximation $A(h)$ to some desired quantity $\backslash cA$. For example, $\backslash cA$ might be the value of some integral $\int_a^b f(x) dx$. For the trapezoidal rule with n steps, $\Delta x = \frac{b-a}{n}$ plays the role of the step size. Often the order of the error generated by the procedure is known. This means

$$\backslash cA = A(h) + Kh^k + K_1h^{k+1} + K_2h^{k+2} + \dots$$

with k being some known constant, called the order of the error, and K, K_1, K_2, \dots being some other (usually unknown) constants. If $A(h)$ is the approximation to $\backslash cA = \int_a^b f(x) dx$ produced by the trapezoidal rule with $\Delta x = h$, then $k = 2$. If Simpson's rule is used, $k = 4$.

Let's first suppose that h is small enough that the terms $K'h^{k+1} + K''h^{k+2} + \dots$ in (E1) are small enough¹ Typically, we don't have access to, and don't care about, the exact error. We only care about its order of magnitude. So if h is small enough that $K_1h^{k+1} + K_2h^{k+2} + \dots$ is a factor of at least, for example, one hundred smaller than Kh^k , then dropping $K_1h^{k+1} + K_2h^{k+2} + \dots$ would not bother us at all. that dropping them has essentially no impact. This would give

$$\backslash cA = A(h) + Kh^k$$

Imagine that we know k , but that we do not know A or K , and think of (E2) as an equation that the unknowns $\backslash cA$ and K have to solve. It may look like we have one equation in the two unknowns $K, \backslash cA$, but that is *not* the case. The reason is that (E2) is (essentially) true for all (sufficiently small) choices of h . If we pick some h , say h_1 , and use the algorithm to determine $A(h_1)$ then (E2), with h replaced by h_1 , gives one equation in the two unknowns $\backslash cA$ and K , and if we then pick some different h , say h_2 , and use the algorithm a second time to determine $A(h_2)$ then (E2), with h replaced by h_2 , gives a second equation in the two unknowns $\backslash cA$ and K . The two equations will then determine both $\backslash cA$ and K .

To be more concrete, suppose that we have picked some specific value of h , and have chosen $h_1 = h$ and $h_2 = \frac{h}{2}$, and that we have evaluated $A(h)$ and $A(h/2)$. Then the two equations are

$$\backslash cA = A(h) + Kh^k \tag{E3a}$$

$$\backslash cA = A(h/2) + K\left(\frac{h}{2}\right)^k \tag{E3b}$$

It is now easy to solve for both K and $\backslash cA$. To get K , just subtract (E3b) from (E3a).

$$\begin{aligned} \text{(E3a)} - \text{(E3b)}: \quad 0 &= A(h) - A(h/2) + \left(1 - \frac{1}{2^k}\right)Kh^k \\ \implies K &= \frac{A(h/2) - A(h)}{[1 - 2^{-k}]h^k} \end{aligned} \tag{E4a}$$

To get $\backslash cA$ multiply (E3b) by 2^k and then subtract (E3a).

$$\begin{aligned} 2^k \text{(E3b)} - \text{(E3a)}: \quad [2^k - 1]\backslash cA &= 2^k A(h/2) - A(h) \\ \implies \backslash cA &= \frac{2^k A(h/2) - A(h)}{2^k - 1} \end{aligned} \tag{E4b}$$

The generation of a "new improved" approximation for $\backslash cA$ from two $A(h)$'s with different values of h is called Richardson² Richardson extrapolation was introduced by the Englishman Lewis Fry Richardson (1881--1953) in 1911. Extrapolation. Here is a summary

Equation C.1.1 Richardson extrapolation

Let $A(h)$ be a step size h approximation to $\backslash cA$. If

$$\backslash cA = A(h) + Kh^k$$

then

$$K = \frac{A(h/2) - A(h)}{[1 - 2^{-k}]h^k} \quad \backslash \text{cA} = \frac{2^k A(h/2) - A(h)}{2^k - 1}$$

This works very well since, by computing $A(h)$ for two different h 's, we can remove the biggest error term in (E1), and so get a much more precise approximation to $\backslash \text{cA}$ for little additional work.

✓ Example C.1.2 Richardson extrapolation with the trapezoidal rule

Applying the trapezoidal rule (1.11.6) to the integral $\backslash \text{cA} = \int_0^1 \frac{4}{1+x^2} dx$ with step sizes $\frac{1}{8}$ and $\frac{1}{16}$ (i.e. with $n = 8$ and $n = 16$) gives, with $h = \frac{1}{8}$,

$$A(h) = 3.1389884945 \quad A(h/2) = 3.1409416120$$

So (E4b), with $k = 2$, gives us the “new improved” approximation

$$\frac{2^2 \times 3.1409416120 - 3.1389884945}{2^2 - 1} = 3.1415926512$$

We saw in Example 1.11.3 that $\int_0^1 \frac{4}{1+x^2} dx = \pi$, so this new approximation really is “improved”:

- $A(1/8)$ agrees with π to two decimal places,
- $A(1/16)$ agrees with π to three decimal places and
- the new approximation agrees with π to eight decimal places.

Beware that (E3b), namely $\backslash \text{cA} = A(h/2) + K(\frac{h}{2})^k$, is saying that $K(\frac{h}{2})^k$ is (approximately) the error in $A(h/2)$, *not* the error in $\backslash \text{cA}$. You cannot get an “even more improved” approximation by using (E4a) to compute K and then adding $K(\frac{h}{2})^k$ to the “new improved” $\backslash \text{cA}$ of (E4b) — doing so just gives $\backslash \text{cA} + K(\frac{h}{2})^k$, not a more accurate $\backslash \text{cA}$.

✓ Example C.1.3 Example 1.11.16 revisited

Suppose again that we wish to use Simpson's rule (1.11.9) to evaluate $\int_0^1 e^{-x^2} dx$ to within an accuracy of 10^{-6} , but that we do not need the degree of certainty provided by Example 1.11.16. Observe that we need (approximately) that $|K|h^4 < 10^{-6}$, so if we can estimate K (using our Richardson trick) then we can estimate the required h . A commonly used strategy, based on this observation, is to

- first apply Simpson's rule twice with some relatively small number of steps and
- then use (E4a), with $k = 4$, to estimate K and
- then use the condition $|K|h^4 \leq 10^{-6}$ to determine, approximately, the number of steps required
- and finally apply Simpson's rule with the number of steps just determined.

Let's implement this strategy. First we estimate K by applying Simpson's rule with step sizes $\frac{1}{4}$ and $\frac{1}{8}$. Writing $\frac{1}{4} = h'$, we get

$$A(h') = 0.74685538 \tag{C.1.1}$$

so that (E4a), with $k = 4$ and h replaced by h' , yields

$$K = \frac{0.74682612 - 0.74685538}{[1 - 2^{-4}](1/4)^4} \tag{C.1.2}$$

We want to use a step size h obeying

$$|K|h^4 \leq 10^{-6} \iff 7.990 \times 10^{-3} h^4 \leq 10^{-6} \iff h \leq \sqrt[4]{\frac{1}{7990}} = \frac{1}{9.45}$$

like, for example, $h = \frac{1}{10}$. Applying Simpson's rule with $h = \frac{1}{10}$ gives

$$A(1/10) = 0.74682495$$

The exact answer, to eight decimal places, is 0.74682413 so the error in $A(1/10)$ is indeed just under 10^{-6} .

Suppose now that we change our minds. We want an accuracy of 10^{-12} , rather than 10^{-6} . We have already estimated K . So now we want to use a step size h obeying

$$\begin{aligned} |K|h^4 \leq 10^{-12} &\iff 7.99 \times 10^{-3} h^4 \leq 10^{-12} \\ &\iff h \leq \sqrt[4]{\frac{1}{7.99 \times 10^9}} = \frac{1}{299.0} \end{aligned}$$

like, for example, $h = \frac{1}{300}$. Applying Simpson's rule with $h = \frac{1}{300}$ gives, to fourteen decimal places,

$$A(1/300) = 0.74682413281344$$

The exact answer, to fourteen decimal places, is 0.74682413281243 so the error in $A(1/300)$ is indeed just over 10^{-12} .

This page titled [C.1: Richardson Extrapolation](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

C.2: Romberg Integration

The formulae (E4a,b) for K and $\backslash cA$ are, of course, only¹“Only” is a bit strong. Don't underestimate the power of a good approximation (pun intended). approximate since they are based on (E2), which is an approximation to (E1). Let's repeat the derivation that leads to (E4), but using the full (E1),

$$\backslash cA = A(h) + Kh^k + K_1h^{k+1} + K_2h^{k+2} + \dots$$

Once again, suppose that we have chosen some h and that we have evaluated $A(h)$ and $A(h/2)$. They obey

$$\backslash cA = A(h) + Kh^k + K_1h^{k+1} + K_2h^{k+2} + \dots \quad (\text{E5a})$$

$$\backslash cA = A(h/2) + K\left(\frac{h}{2}\right)^k + K_1\left(\frac{h}{2}\right)^{k+1} + K_2\left(\frac{h}{2}\right)^{k+2} + \dots \quad (\text{E5b})$$

Now, as we did in the derivation of (E4b), multiply (E5b) by 2^k and then subtract (E5a). This gives

$$(2^k - 1) \backslash cA = 2^k A(h/2) - A(h) - \backslash half K_1 h^{k+1} - \frac{3}{4} K_2 h^{k+2} + \dots$$

and then, dividing across by $(2^k - 1)$,

$$\backslash cA = \frac{2^k A(h/2) - A(h)}{2^k - 1} - \frac{1/2}{2^k - 1} K_1 h^{k+1} - \frac{3/4}{2^k - 1} K_2 h^{k+2} + \dots$$

Hence if we define our “new improved approximation”

$$B(h) = \frac{2^k A(h/2) - A(h)}{2^k - 1} \text{ and } \tilde{K} = -\frac{1/2}{2^k - 1} K_1 \text{ and } \tilde{K}_1 = -\frac{3/4}{2^k - 1} K_2$$

we have

$$\backslash cA = B(h) + \tilde{K}h^{k+1} + \tilde{K}_1h^{k+2} + \dots$$

which says that $B(h)$ is an approximation to $\backslash cA$ whose error is of order $k+1$, one better²That is, the error decays as h^{k+1} as opposed to h^k — so, as h decreases, it gets smaller faster. than $A(h)$'s.

If $A(h)$ has been computed for three values of h , we can generate $B(h)$ for two values of h and repeat the above procedure with a new value of k . And so on. One widely used numerical integration algorithm, called Romberg integration³Romberg Integration was introduced by the German Werner Romberg (1909–2003) in 1955., applies this procedure repeatedly to the trapezoidal rule. It is known that the trapezoidal rule approximation $T(h)$ to an integral I has error behaviour (assuming that the integrand $f(x)$ is smooth)

$$I = T(h) + K_1h^2 + K_2h^4 + K_3h^6 + \dots$$

Only even powers of h appear. Hence

$$T(h) \quad \text{has error of order } 2$$

so that, using (E6) with $k = 2$,

$$T_1(h) = \frac{4T(h/2) - T(h)}{3} \quad \text{has error of order } 4$$

so that, using (E6) with $k = 4$,

$$T_2(h) = \frac{16T_1(h/2) - T_1(h)}{15} \quad \text{has error of order } 6$$

so that, using (E6) with $k = 6$,

$$T_3(h) = \frac{64T_2(h/2) - T_2(h)}{63} \quad \text{has error of order } 8$$

and so on. We know another method which produces an error of order 4 — Simpson's rule. In fact, $T_1(h)$ is exactly Simpson's rule (for step size $\frac{h}{2}$).

Equation C.2.1 Romberg integration

Let $T(h)$ be the trapezoidal rule approximation, with step size h , to an integral I . The Romberg integration algorithm is

$$\begin{aligned}
 T_1(h) &= \frac{4T(h/2) - T(h)}{3} \\
 T_2(h) &= \frac{16T_1(h/2) - T_1(h)}{15} \\
 T_3(h) &= \frac{64T_2(h/2) - T_2(h)}{63} \\
 &\vdots \\
 T_k(h) &= \frac{2^{2k}T_{k-1}(h/2) - T_{k-1}(h)}{2^{2k} - 1} \\
 &\vdots
 \end{aligned}$$

Example C.2.2 Finding π by Romberg integration

The following table⁴ illustrates Romberg integration by applying it to the area A of the integral $A = \int_0^1 \frac{4}{1+x^2} dx$. The second column, for example, of the table only reports 5 decimal places for $T(h)$. But many more decimal places of $T(h)$ were used in the computations of $T_1(h)$ etc. The exact value of this integral is π which is 3.14159265358979 to fourteen decimal places.

h	$T(h)$	$T_1(h)$	$T_2(h)$	$T_3(h)$
1/4	3.13118	3.14159250246	3.14159266114	3.14159265359003
1/8	3.13899	3.141592651225	3.141592653708	
1/16	3.14094	3.141592653553		
1/32	3.14143			

This computation required the evaluation of $f(x) = \frac{4}{1+x^2}$ only for $x = \frac{n}{32}$ with $0 \leq n \leq 32$ — that is, a total of 33 evaluations of f . Those 33 evaluations gave us 12 correct decimal places. By way of comparison, $T(\frac{1}{32})$ used the same 33 evaluations of f , but only gave us 3 correct decimal places.

As we have seen, Richardson extrapolation can be used to choose the step size so as to achieve some desired degree of accuracy. We are next going to consider a family of algorithms that extend this idea to use small step sizes in the part of the domain of integration where it is hard to get good accuracy and large step sizes in the part of the domain of integration where it is easy to get good accuracy. We will illustrate the ideas by applying them to the integral $\int_0^1 \sqrt{x} dx$. The integrand \sqrt{x} changes very quickly when x is small and changes slowly when x is large. So we will make the step size small near $x = 0$ and make the step size large near $x = 1$.

This page titled [C.2: Romberg Integration](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

C.3: Adaptive Quadrature

Richardson extrapolation is also used to choose the step size so as to achieve some desired degree of accuracy. “Adaptive quadrature” refers to a family of algorithms that use small step sizes in the part of the domain of integration where it is hard to get good accuracy and large step sizes in the part of the domain of integration where it is easy to get good accuracy.

We'll illustrate the idea using Simpson's rule applied to the integral $\int_a^b f(x) dx$, and assuming that we want the error to be no more than (approximately) some fixed constant ε . For example, ε could be 10^{-6} . Denote by $S(a', b'; h')$, the answer given when Simpson's rule is applied to the integral $\int_{a'}^{b'} f(x) dx$ with step size h' .

- *Step 1.* We start by applying Simpson's rule, combined with Richardson extrapolation so as to get an error estimate, with the largest possible step size h . Namely, set $h = \frac{b-a}{2}$ and compute

$$f(a) \quad f\left(a + \frac{h}{2}\right) \quad f(a+h) = f\left(\frac{a+b}{2}\right) \quad f\left(a + \frac{3h}{2}\right) \quad f(a+2h) = f(b)$$

Then

$$S(a, b; h) = \frac{h}{3} \{f(a) + 4f(a+h) + f(b)\}$$

and

$$\begin{aligned} S\left(a, b; \frac{h}{2}\right) &= \frac{h}{6} \{f(a) + 4f\left(a + \frac{h}{2}\right) + 2f\left(\frac{a+b}{2}\right) + 4f\left(a + \frac{3h}{2}\right) + f(b)\} \\ &= S\left(a, \frac{a+b}{2}; \frac{h}{2}\right) + S\left(\frac{a+b}{2}, b; \frac{h}{2}\right) \end{aligned}$$

with

$$\begin{aligned} S\left(a, \frac{a+b}{2}; \frac{h}{2}\right) &= \frac{h}{6} \{f(a) + 4f\left(a + \frac{h}{2}\right) + f\left(\frac{a+b}{2}\right)\} \\ S\left(\frac{a+b}{2}, b; \frac{h}{2}\right) &= \frac{h}{6} \{f\left(\frac{a+b}{2}\right) + 4f\left(a + \frac{3h}{2}\right) + f(b)\} \end{aligned}$$

Using the Richardson extrapolation formula (E4a) with $k = 4$ gives that the error in $S(a, b; \frac{h}{2})$ is (approximately)

$$\left|K\left(\frac{h}{2}\right)^4\right| = \frac{1}{15} \left|S\left(a, b; \frac{h}{2}\right) - S(a, b; h)\right|$$

If this is smaller than ε , we have (approximately) the desired accuracy and stop¹ It is very common to build in a bit of a safety margin and require that, for example, $|K(\frac{h}{2})^4|$ be smaller than $\frac{\varepsilon}{2}$ rather than ε .

- *Step 2.* If (E7) is larger than ε , we divide the original integral $I = \int_a^b f(x) dx$ into two “half-sized” integrals, $I_1 = \int_a^{\frac{a+b}{2}} f(x) dx$ and $I_2 = \int_{\frac{a+b}{2}}^b f(x) dx$ and repeat the procedure of Step 1 on each of them, but with h replaced by $\frac{h}{2}$ and ε replaced by $\frac{\varepsilon}{2}$ — if we can find an approximation, \tilde{I}_1 , to I_1 with an error less than $\frac{\varepsilon}{2}$ and an approximation, \tilde{I}_2 , to I_2 with an error less than $\frac{\varepsilon}{2}$, then $\tilde{I}_1 + \tilde{I}_2$ approximates I with an error less than ε . Here is more detail.
 - If the error in the approximation \tilde{I}_1 to I_1 and the error in the approximation \tilde{I}_2 to I_2 are both acceptable, then we use \tilde{I}_1 as our final approximation to I_1 and we use \tilde{I}_2 as our final approximation to I_2 .
 - If the error in the approximation \tilde{I}_1 to I_1 is acceptable but the error in the approximation \tilde{I}_2 to I_2 is not acceptable, then we use \tilde{I}_1 as our final approximation to I_1 but we subdivide the integral I_2 .
 - If the error in the approximation \tilde{I}_1 to I_1 is not acceptable but the error in the approximation \tilde{I}_2 to I_2 is acceptable, then we use \tilde{I}_2 as our final approximation to I_2 but we subdivide the integral I_1 .
 - If the error in the approximation \tilde{I}_1 to I_1 and the error in the approximation \tilde{I}_2 to I_2 are both not acceptable, then we subdivide both of the integrals I_1 and I_2 .

So we *adapt* the step size as we go.

- *Steps 3, 4, 5, ...* Repeat as required.

✓ Example C.3.1 Adaptive quadrature

Let's apply adaptive quadrature using Simpson's rule as above with the goal of computing $\int_0^1 \sqrt{x} \, dx$ with an error of at most $\varepsilon = 0.0005 = 5 \times 10^{-4}$. Observe that $\frac{d}{dx} \sqrt{x} = \frac{1}{2\sqrt{x}}$ blows up as x tends to zero. The integrand changes very quickly when x is small. So we will probably need to make the step size small near the limit of integration $x = 0$.

- *Step 1 — the interval $[0, 1]$.* (The notation $[0, 1]$ stands for the interval $0 \leq x \leq 1$.)

$$S(0, 1; \frac{1}{2}) = 0.63807119$$

$$S(0, \frac{1}{2}; \frac{1}{4}) = 0.22559223$$

$$S(\frac{1}{2}, 1; \frac{1}{4}) = 0.43093403$$

$$\begin{aligned} \text{error} &= \frac{1}{15} |S(0, \frac{1}{2}; \frac{1}{4}) + S(\frac{1}{2}, 1; \frac{1}{4}) - S(0, 1; \frac{1}{2})| = 0.0012 \\ &> \varepsilon = 0.0005 \end{aligned}$$

This is unacceptably large, so we subdivide the interval $[0, 1]$ into the two halves $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$ and apply the procedure separately to each half.

- *Step 2a — the interval $[0, \frac{1}{2}]$.*

$$S(0, \frac{1}{2}; \frac{1}{4}) = 0.22559223$$

$$S(0, \frac{1}{4}; \frac{1}{8}) = 0.07975890$$

$$S(\frac{1}{4}, \frac{1}{2}; \frac{1}{8}) = 0.15235819$$

$$\begin{aligned} \text{error} &= \frac{1}{15} |S(0, \frac{1}{4}; \frac{1}{8}) + S(\frac{1}{4}, \frac{1}{2}; \frac{1}{8}) - S(0, \frac{1}{2}; \frac{1}{4})| = 0.00043 \\ &> \frac{\varepsilon}{2} = 0.00025 \end{aligned}$$

This error is unacceptably large.

- *Step 2b — the interval $[\frac{1}{2}, 1]$.*

$$S(\frac{1}{2}, 1; \frac{1}{4}) = 0.43093403$$

$$S(\frac{1}{2}, \frac{3}{4}; \frac{1}{8}) = 0.19730874$$

$$S(\frac{3}{4}, 1; \frac{1}{8}) = 0.23365345$$

$$\begin{aligned} \text{error} &= \frac{1}{15} |S(\frac{1}{2}, \frac{3}{4}; \frac{1}{8}) + S(\frac{3}{4}, 1; \frac{1}{8}) - S(\frac{1}{2}, 1; \frac{1}{4})| = 0.0000019 \\ &< \frac{\varepsilon}{2} = 0.00025 \end{aligned}$$

This error is acceptable.

- *Step 2 resumé.* The error for the interval $[\frac{1}{2}, 1]$ is small enough, so we accept

$$S(\frac{1}{2}, 1; \frac{1}{8}) = S(\frac{1}{2}, \frac{3}{4}; \frac{1}{8}) + S(\frac{3}{4}, 1; \frac{1}{8}) = 0.43096219$$

as the approximate value of $\int_{1/2}^1 \sqrt{x} \, dx$.

The error for the interval $[0, \frac{1}{2}]$ is unacceptably large, so we subdivide the interval $[0, \frac{1}{2}]$ into the two halves $[0, \frac{1}{4}]$ and $[\frac{1}{4}, \frac{1}{2}]$ and apply the procedure separately to each half.

- *Step 3a — the interval $[0, \frac{1}{4}]$.*

$$S(0, \frac{1}{4}; \frac{1}{8}) = 0.07975890$$

$$S(0, \frac{1}{8}; \frac{1}{16}) = 0.02819903$$

$$S(\frac{1}{8}, \frac{1}{4}; \frac{1}{16}) = 0.05386675$$

$$\begin{aligned} \text{error} &= \frac{1}{15} |S(0, \frac{1}{8}; \frac{1}{16}) + S(\frac{1}{8}, \frac{1}{4}; \frac{1}{16}) - S(0, \frac{1}{4}; \frac{1}{8})| \\ &= 0.000153792 > \frac{\varepsilon}{4} = 0.000125 \end{aligned}$$

This error is unacceptably large.

- *Step 3b — the interval $[\frac{1}{4}, \frac{1}{2}]$.*

$$\begin{aligned}
 S\left(\frac{1}{4}, \frac{1}{2}; \frac{1}{8}\right) &= 0.15235819 \\
 S\left(\frac{1}{4}, \frac{3}{8}; \frac{1}{16}\right) &= 0.06975918 \\
 S\left(\frac{3}{8}, \frac{1}{2}; \frac{1}{16}\right) &= 0.08260897 \\
 \text{error} &= \frac{1}{15} \left| S\left(\frac{1}{4}, \frac{3}{8}; \frac{1}{16}\right) + S\left(\frac{3}{8}, \frac{1}{2}; \frac{1}{16}\right) - S\left(\frac{1}{4}, \frac{1}{2}; \frac{1}{8}\right) \right| \\
 &= 0.00000066 < \frac{\varepsilon}{4} = 0.000125
 \end{aligned}$$

This error is acceptable.

- *Step 3 resumé.* The error for the interval $[\frac{1}{4}, \frac{1}{2}]$ is small enough, so we accept

$$S\left(\frac{1}{4}, \frac{1}{2}; \frac{1}{16}\right) = S\left(\frac{1}{4}, \frac{3}{8}; \frac{1}{16}\right) + S\left(\frac{3}{8}, \frac{1}{2}; \frac{1}{16}\right) = 0.15236814$$

as the approximate value of $\int_{1/4}^{1/2} \sqrt{x} \, dx$.

The error for the interval $[0, \frac{1}{4}]$ is unacceptably large, so we subdivide the interval $[0, \frac{1}{4}]$ into the two halves $[0, \frac{1}{8}]$ and $[\frac{1}{8}, \frac{1}{4}]$ and apply the procedure separately to each half.

- *Step 4a — the interval $[0, \frac{1}{8}]$.*

$$\begin{aligned}
 S\left(0, \frac{1}{8}; \frac{1}{16}\right) &= 0.02819903 \\
 S\left(0, \frac{1}{16}; \frac{1}{32}\right) &= 0.00996986 \\
 S\left(\frac{1}{16}, \frac{1}{8}; \frac{1}{32}\right) &= 0.01904477 \\
 \text{error} &= \frac{1}{15} \left| S\left(0, \frac{1}{16}; \frac{1}{32}\right) + S\left(\frac{1}{16}, \frac{1}{8}; \frac{1}{32}\right) - S\left(0, \frac{1}{8}; \frac{1}{16}\right) \right| \\
 &= 0.000054 < \frac{\varepsilon}{8} = 0.0000625
 \end{aligned}$$

This error is acceptable.

- *Step 4b — the interval $[\frac{1}{8}, \frac{1}{4}]$.*

$$\begin{aligned}
 S\left(\frac{1}{8}, \frac{1}{4}; \frac{1}{16}\right) &= 0.05386675 \\
 S\left(\frac{1}{8}, \frac{3}{16}; \frac{1}{32}\right) &= 0.02466359 \\
 S\left(\frac{3}{16}, \frac{1}{4}; \frac{1}{32}\right) &= 0.02920668 \\
 \text{error} &= \frac{1}{15} \left| S\left(\frac{1}{8}, \frac{3}{16}; \frac{1}{32}\right) + S\left(\frac{3}{16}, \frac{1}{4}; \frac{1}{32}\right) - S\left(\frac{1}{8}, \frac{1}{4}; \frac{1}{16}\right) \right| \\
 &= 0.00000024 < \frac{\varepsilon}{8} = 0.0000625
 \end{aligned}$$

This error is acceptable.

- *Step 4 resumé.* The error for the interval $[0, \frac{1}{8}]$ is small enough, so we accept

$$S\left(0, \frac{1}{8}; \frac{1}{32}\right) = S\left(0, \frac{1}{16}; \frac{1}{32}\right) + S\left(\frac{1}{16}, \frac{1}{8}; \frac{1}{32}\right) = 0.02901464$$

as the approximate value of $\int_0^{1/8} \sqrt{x} \, dx$.

The error for the interval $[\frac{1}{8}, \frac{1}{4}]$ is small enough, so we accept

$$S\left(\frac{1}{8}, \frac{1}{4}; \frac{1}{32}\right) = S\left(\frac{1}{8}, \frac{3}{16}; \frac{1}{32}\right) + S\left(\frac{3}{16}, \frac{1}{4}; \frac{1}{32}\right) = 0.05387027$$

as the approximate value of $\int_{1/8}^{1/4} \sqrt{x} \, dx$.

- *Conclusion.* The approximate value for $\int_0^1 \sqrt{x} \, dx$ is

$$\begin{aligned}
 S\left(0, \frac{1}{8}; \frac{1}{32}\right) + S\left(\frac{1}{8}, \frac{1}{4}; \frac{1}{32}\right) + S\left(\frac{1}{4}, \frac{1}{2}; \frac{1}{16}\right) + S\left(\frac{1}{2}, 1; \frac{1}{8}\right) \\
 = 0.66621525
 \end{aligned} \tag{E8}$$

Of course the exact value of $\int_0^1 \sqrt{x} \, dx = \frac{2}{3}$, so the actual error in our approximation is

$$\frac{2}{3} - 0.66621525 = 0.00045 < \varepsilon = 0.0005$$

Here is what Simpson's rule gives us when applied with some fixed step sizes.

$$S(0, 1; \frac{1}{8}) = 0.66307928$$

$$S(0, 1; \frac{1}{16}) = 0.66539819$$

$$S(0, 1; \frac{1}{32}) = 0.66621818$$

$$S(0, 1; \frac{1}{64}) = 0.66650810$$

So to get an error comparable to that in (E8) from Simpson's rule with a fixed step size, we need to use $h = \frac{1}{32}$. In (E8) the step size $h = \frac{1}{32}$ was just used on the subinterval $[0, \frac{1}{4}]$.

This page titled [C.3: Adaptive Quadrature](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

D: Numerical Solution of ODE's

In Section 2.4 we solved a number of initial value problems of the form

$$\begin{aligned}y'(t) &= f(t, y(t)) \\ y(t_0) &= y_0\end{aligned}$$

Here $f(t, y)$ is a given function, t_0 is a given initial time and y_0 is a given initial value for y . The unknown in the problem is the function $y(t)$. There are a number of other techniques for analytically solving some problems of this type. However it is often simply not possible to find an explicit solution. This appendix introduces some simple algorithms for generating approximate numerical solutions to such problems.

This page titled [D: Numerical Solution of ODE's](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

D.1: Simple ODE Solvers — Derivation

The first order of business is to derive three simple algorithms for generating approximate numerical solutions to the initial value problem

$$\begin{aligned}y'(t) &= f(t, y(t)) \\ y(t_0) &= y_0\end{aligned}$$

The first is called Euler's method because it was developed by (surprise!) Euler¹.

Euler's Method

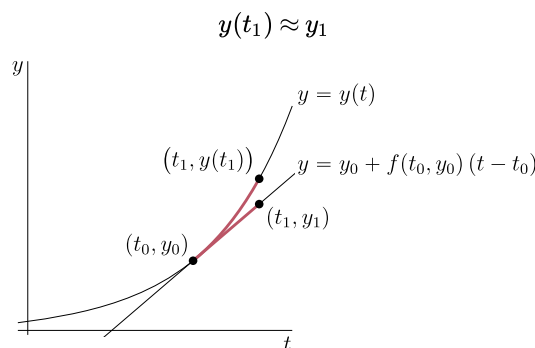
Our goal is to approximate (numerically) the unknown function

$$\begin{aligned}y(t) &= y(t_0) + \int_{t_0}^t y'(\tau) d\tau \\ &= y(t_0) + \int_{t_0}^t f(\tau, y(\tau)) d\tau\end{aligned}$$

for $t \geq t_0$. We are told explicitly the value of $y(t_0)$, namely y_0 . So we know $f(\tau, y(\tau))|_{\tau=t_0} = f(t_0, y_0)$. But we do not know the integrand $f(\tau, y(\tau))$ for $\tau > t_0$. On the other hand, if τ is close t_0 , then $f(\tau, y(\tau))$ will remain close² This will be the case as long as $f(t, y)$ is continuous. to $f(t_0, y_0)$. So pick a small number h and define

$$\begin{aligned}t_1 &= t_0 + h \\ y_1 &= y(t_0) + \int_{t_0}^{t_1} f(t_0, y_0) d\tau = y_0 + f(t_0, y_0)(t_1 - t_0) \\ &= y_0 + f(t_0, y_0)h\end{aligned}$$

By the above argument



Now we start over from the new point (t_1, y_1) . We now know an approximate value for y at time t_1 . If $y(t_1)$ were exactly y_1 , then the instantaneous rate of change of y at time t_1 , namely $y'(t_1) = f(t_1, y(t_1))$, would be exactly $f(t_1, y_1)$ and $f(t, y(t))$ would remain close to $f(t_1, y_1)$ for t close to t_1 . Defining

$$\begin{aligned}t_2 &= t_1 + h = t_0 + 2h \\ y_2 &= y_1 + \int_{t_1}^{t_2} f(t_1, y_1) dt = y_1 + f(t_1, y_1)(t_2 - t_1) \\ &= y_1 + f(t_1, y_1)h\end{aligned}$$

we have

$$y(t_2) \approx y_2$$

We just repeat this argument ad infinitum. Define, for $n = 0, 1, 2, 3, \dots$

$$t_n = t_0 + nh$$

Suppose that, for some value of n , we have already computed an approximate value y_n for $y(t_n)$. Then the rate of change of $y(t)$ for t close to t_n is $f(t, y(t)) \approx f(t_n, y(t_n)) \approx f(t_n, y_n)$ and

Equation D.1.1 Euler's Method

$$y(t_{n+1}) \approx y_{n+1} = y_n + f(t_n, y_n)h$$

This algorithm is called *Euler's Method*. The parameter h is called the *step size*.

Here is a table applying a few steps of Euler's method to the initial value problem

$$\begin{aligned} y' &= -2t + y \\ y(0) &= 3 \end{aligned}$$

with step size $h = 0.1$. For this initial value problem

$$\begin{aligned} f(t, y) &= -2t + y \\ t_0 &= 0 \\ y_0 &= 3 \end{aligned}$$

Of course this initial value problem has been chosen for illustrative purposes only. The exact solution is³ $y(t) = 2 + 2t + e^t$.

n	t_n	y_n	$f(t_n, y_n) = -2t_n + y_n$	$y_{n+1} = y_n + f(t_n, y_n) * h$
0	0.0	3.000	$-2 * 0.0 + 3.000 = 3.000$	$3.000 + 3.000 * 0.1 = 3.300$
1	0.1	3.300	$-2 * 0.1 + 3.300 = 3.100$	$3.300 + 3.100 * 0.1 = 3.610$
2	0.2	3.610	$-2 * 0.2 + 3.610 = 3.210$	$3.610 + 3.210 * 0.1 = 3.931$
3	0.3	3.931	$-2 * 0.3 + 3.931 = 3.331$	$3.931 + 3.331 * 0.1 = 4.264$
4	0.4	4.264	$-2 * 0.4 + 4.264 = 3.464$	$4.264 + 3.464 * 0.1 = 4.611$
5	0.5	4.611		

The exact solution at $t = 0.5$ is 4.6487, to four decimal places. We expect that Euler's method will become more accurate as the step size becomes smaller. But, of course, the amount of effort goes up as well. If we recompute using $h = 0.01$, we get (after much more work) 4.6446.

The Improved Euler's Method

Euler's method is one algorithm which generates approximate solutions to the initial value problem

$$\begin{aligned} y'(t) &= f(t, y(t)) \\ y(t_0) &= y_0 \end{aligned}$$

In applications, $f(t, y)$ is a given function and t_0 and y_0 are given numbers. The function $y(t)$ is unknown. Denote by $\varphi(t)$ the exact solution⁴ Under reasonable hypotheses on f , there is exactly one such solution. The interested reader should search engine their way to the Picard-Lindelöf theorem. for this initial value problem. In other words $\varphi(t)$ is the function that obeys

$$\begin{aligned} \varphi'(t) &= f(t, \varphi(t)) \\ \varphi(t_0) &= y_0 \end{aligned}$$

exactly.

Fix a step size h and define $t_n = t_0 + nh$. By turning the problem into one of approximating integrals, we now derive another algorithm that generates approximate values for φ at the sequence of equally spaced time values t_0, t_1, t_2, \dots . We shall denote the approximate values y_n with

$$y_n \approx \varphi(t_n)$$

By the fundamental theorem of calculus and the differential equation, the exact solution obeys

$$\begin{aligned}\varphi(t_{n+1}) &= \varphi(t_n) + \int_{t_n}^{t_{n+1}} \varphi'(t) dt \\ &= \varphi(t_n) + \int_{t_n}^{t_{n+1}} f(t, \varphi(t)) dt\end{aligned}$$

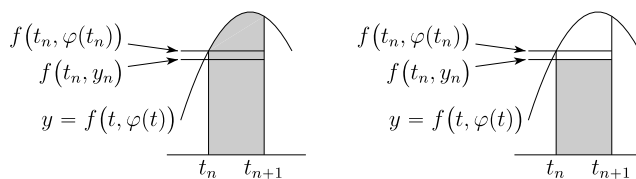
Fix any n and suppose that we have already found y_0, y_1, \dots, y_n . Our algorithm for computing y_{n+1} will be of the form

$$y_{n+1} = y_n + \text{approximate value of } \int_{t_n}^{t_{n+1}} f(t, \varphi(t)) dt$$

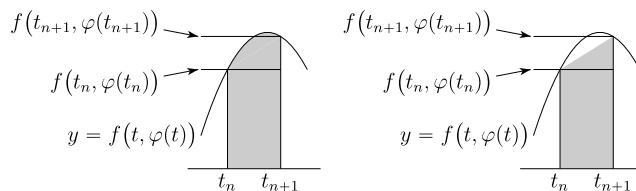
In Euler's method, we approximated $f(t, \varphi(t))$ for $t_n \leq t \leq t_{n+1}$ by the constant $f(t_n, y_n)$. Thus

$$\begin{aligned}\text{Euler's approximate value for } \int_{t_n}^{t_{n+1}} f(t, \varphi(t)) dt &\text{ is} \\ \int_{t_n}^{t_{n+1}} f(t_n, y_n) dt &= f(t_n, y_n)h\end{aligned}$$

So Euler approximates the area of the complicated region $0 \leq y \leq f(t, \varphi(t)), t_n \leq t \leq t_{n+1}$ (represented by the shaded region under the parabola in the left half of the figure below) by the area of the rectangle $0 \leq y \leq f(t_n, y_n), t_n \leq t \leq t_{n+1}$ (the shaded rectangle in the right half of the figure below).



Our second algorithm, the improved Euler's method, gets a better approximation by using the trapezoidal rule. That is, we approximate the integral by the area of the trapezoid on the right below, rather than the rectangle on the right above.



The exact area of this trapezoid is the length h of the base multiplied by the average of the heights of the two sides, which is $\frac{1}{2} [f(t_n, \varphi(t_n)) + f(t_{n+1}, \varphi(t_{n+1}))]$. Of course we do not know $\varphi(t_n)$ or $\varphi(t_{n+1})$ exactly.

Recall that we have already found y_0, \dots, y_n and are in the process of finding y_{n+1} . So we already have an approximation for $\varphi(t_n)$, namely y_n . But we still need to approximate $\varphi(t_{n+1})$. We can do so by using one step of the original Euler method! That is

$$\varphi(t_{n+1}) \approx \varphi(t_n) + \varphi'(t_n)h \approx y_n + f(t_n, y_n)h$$

So our approximation of $\frac{1}{2} [f(t_n, \varphi(t_n)) + f(t_{n+1}, \varphi(t_{n+1}))]$ is

$$\frac{1}{2} [f(t_n, y_n) + f(t_{n+1}, y_n + f(t_n, y_n)h)]$$

and

$$\begin{aligned}\text{Improved Euler's approximate value for } \int_{t_n}^{t_{n+1}} f(t, \varphi(t)) dt &\text{ is} \\ \frac{1}{2} [f(t_n, y_n) + f(t_{n+1}, y_n + f(t_n, y_n)h)] &h\end{aligned}$$

Putting everything together⁵, the improved Euler's method algorithm is

Equation D.1.2 Improved Euler

$$y(t_{n+1}) \approx y_{n+1} = y_n + \frac{1}{2} [f(t_n, y_n) + f(t_{n+1}, y_n + f(t_n, y_n)h)] h$$

Here are the first two steps of the improved Euler's method applied to

$$y' = -2t + y \quad y(0) = 3$$

with $h = 0.1$. In each step we compute $f(t_n, y_n)$, followed by $y_n + f(t_n, y_n)h$, which we denote \tilde{y}_{n+1} , followed by $f(t_{n+1}, \tilde{y}_{n+1})$, followed by $y_{n+1} = y_n + \frac{1}{2} [f(t_n, y_n) + f(t_{n+1}, \tilde{y}_{n+1})] h$.

$$\begin{aligned} t_0 = 0 \quad y_0 = 3 &\implies f(t_0, y_0) = -2 * 0 + 3 = 3 \\ &\implies \tilde{y}_1 = 3 + 3 * 0.1 = 3.3 \\ &\implies f(t_1, \tilde{y}_1) = -2 * 0.1 + 3.3 = 3.1 \\ &\implies y_1 = 3 + \frac{1}{2} [3 + 3.1] * 0.1 = 3.305 \\ t_1 = 0.1 \quad y_1 = 3.305 &\implies f(t_1, y_1) = -2 * 0.1 + 3.305 = 3.105 \\ &\implies \tilde{y}_2 = 3.305 + 3.105 * 0.1 = 3.6155 \\ &\implies f(t_2, \tilde{y}_2) = -2 * 0.2 + 3.6155 = 3.2155 \\ &\implies y_2 = 3.305 + \frac{1}{2} [3.105 + 3.2155] * 0.1 \\ &\implies = 3.621025 \end{aligned}$$

Here is a table which gives the first five steps.

n	t_n	y_n	$f(t_n, y_n)$	\tilde{y}_{n+1}	$f(t_{n+1}, \tilde{y}_{n+1})$	y_{n+1}
0	0.0	3.000	3.000	3.300	3.100	3.305
1	0.1	3.305	3.105	3.616	3.216	3.621
2	0.2	3.621	3.221	3.943	3.343	3.949
3	0.3	3.949	3.349	4.284	3.484	4.291
4	0.4	4.291	3.491	4.640	3.640	4.647
5	0.5	4.647				

As we saw at the end of Section D.1.1, the exact $y(0.5)$ is 4.6487, to four decimal places, and Euler's method gave 4.611.

The Runge-Kutta Method

The Runge-Kutta⁶ algorithm is similar to the Euler and improved Euler methods in that it also uses, in the notation of the last subsection,

$$y_{n+1} = y_n + \text{approximate value for } \int_{t_n}^{t_{n+1}} f(t, \varphi(t)) dt$$

But rather than approximating $\int_{t_n}^{t_{n+1}} f(t, \varphi(t)) dt$ by the area of a rectangle, as does Euler, or by the area of a trapezoid, as does improved Euler, it approximates by the area under a parabola. That is, it uses Simpson's rule. According to Simpson's rule (which is derived in Section 1.11.3)

$$\begin{aligned} \int_{t_n}^{t_n+h} f(t, \varphi(t)) dt \\ \approx \frac{h}{6} \left[f(t_n, \varphi(t_n)) + 4f\left(t_n + \frac{h}{2}, \varphi\left(t_n + \frac{h}{2}\right)\right) + f\left(t_n + h, \varphi\left(t_n + h\right)\right) \right] \end{aligned}$$

Analogously to what happened in our development of the improved Euler method, we don't know $\varphi(t_n)$, $\varphi\left(t_n + \frac{h}{2}\right)$ or $\varphi(t_n + h)$. So we have to approximate them as well. The Runge-Kutta algorithm, incorporating all these approximations, is⁷ It is well beyond

our scope to derive this algorithm, though the derivation is similar in flavour to that of the improved Euler method. You can find more in, for example, Wikipedia.

Equation D.1.3 Runge-Kutta

$$\begin{aligned}
 k_{1,n} &= f(t_n, y_n) \\
 k_{2,n} &= f\left(t_n + \frac{1}{2}h, y_n + \frac{h}{2}k_{1,n}\right) \\
 k_{3,n} &= f\left(t_n + \frac{1}{2}h, y_n + \frac{h}{2}k_{2,n}\right) \\
 k_{4,n} &= f(t_n + h, y_n + hk_{3,n}) \\
 y_{n+1} &= y_n + \frac{h}{6}[k_{1,n} + 2k_{2,n} + 2k_{3,n} + k_{4,n}]
 \end{aligned}$$

That is, Runge-Kutta uses

- $k_{1,n}$ to approximate $f(t_n, \varphi(t_n)) = \varphi'(t_n)$,
- both $k_{2,n}$ and $k_{3,n}$ to approximate $f\left(t_n + \frac{h}{2}, \varphi\left(t_n + \frac{h}{2}\right)\right) = \varphi'\left(t_n + \frac{h}{2}\right)$, and
- $k_{4,n}$ to approximate $f(t_n + h, \varphi(t_n + h))$.

Here are the first two steps of the Runge-Kutta algorithm applied to

$$y' = -2t + y \quad y(0) = 3$$

with $h = 0.1$.

$$\begin{aligned}
 t_0 = 0 \quad y_0 = 3 \\
 \implies k_{1,0} &= f(0, 3) = -2 * 0 + 3 = 3 \\
 \implies y_0 + \frac{h}{2}k_{1,0} &= 3 + 0.05 * 3 = 3.15 \\
 \implies k_{2,0} &= f(0.05, 3.15) = -2 * 0.05 + 3.15 = 3.05 \\
 \implies y_0 + \frac{h}{2}k_{2,0} &= 3 + 0.05 * 3.05 = 3.1525 \\
 \implies k_{3,0} &= f(0.05, 3.1525) = -2 * 0.05 + 3.1525 = 3.0525 \\
 \implies y_0 + hk_{3,0} &= 3 + 0.1 * 3.0525 = 3.30525 \\
 \implies k_{4,0} &= f(0.1, 3.30525) = -2 * 0.1 + 3.30525 = 3.10525 \\
 \implies y_1 &= 3 + \frac{0.1}{6}[3 + 2 * 3.05 + 2 * 3.0525 + 3.10525] = 3.3051708 \\
 t_1 = 0.1 \quad y_1 = 3.3051708 \\
 \implies k_{1,1} &= f(0.1, 3.3051708) = -2 * 0.1 + 3.3051708 = 3.1051708 \\
 \implies y_1 + \frac{h}{2}k_{1,1} &= 3.3051708 + 0.05 * 3.1051708 = 3.4604293 \\
 \implies k_{2,1} &= f(0.15, 3.4604293) = -2 * 0.15 + 3.4604293 = 3.1604293 \\
 \implies y_1 + \frac{h}{2}k_{2,1} &= 3.3051708 + 0.05 * 3.1604293 = 3.4631923 \\
 \implies k_{3,1} &= f(0.15, 3.4631923) = -2 * 0.15 + 3.4631923 = 3.1631923 \\
 \implies y_1 + hk_{3,1} &= 3.3051708 + 0.1 * 3.4631923 = 3.62149 \\
 \implies k_{4,1} &= f(0.2, 3.62149) = -2 * 0.2 + 3.62149 = 3.22149 \\
 \implies y_2 &= 3.3051708 + \frac{0.1}{6}[3.1051708 + 2 * 3.1604293 + \\
 &\quad + 2 * 3.1631923 + 3.22149] = 3.6214025 \\
 t_2 = 0.2 \quad y_2 = 3.6214025
 \end{aligned}$$

Now, while this might look intimidating written out in full like this, one should keep in mind that it is quite easy to write a program to do this. Here is a table giving the first five steps. The data is only given to three decimal places even though the computation has been done to many more.

n	t_n	y_n	$k_{1,n}$	$y_{n,1}$	$k_{2,n}$	$y_{n,2}$	$k_{3,n}$	$y_{n,3}$	$k_{4,n}$	y_{n+1}
0	0.0	3.000	3.000	3.150	3.050	3.153	3.053	3.305	3.105	3.305
1	0.1	3.305	3.105	3.460	3.160	3.463	3.163	3.621	3.221	3.621

2	0.2	3.621	3.221	3.782	3.282	3.786	3.286	3.950	3.350	3.949
3	0.3	3.950	3.350	4.117	3.417	4.121	3.421	4.292	3.492	4.291
4	0.4	4.292	3.492	4.466	3.566	4.470	3.570	4.649	3.649	4.648
5	0.5	4.648720 6								

As we saw at the end of Section D.1.2, the exact $y(0.5)$ is 4.6487213, to seven decimal places, Euler's method gave 4.611, and improved Euler gave 4.647.

So far we have, hopefully, motivated the Euler, improved Euler and Runge-Kutta algorithms. We have not attempted to see how efficient and how accurate the algorithms are. A first look at those questions is provided in the next section.

This page titled D.1: Simple ODE Solvers — Derivation is shared under a [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

D.2: Simple ODE Solvers — Error Behaviour

We now provide an introduction to the error behaviour of Euler's Method, the improved Euler's method and the Runge-Kutta algorithm for generating approximate solutions to the initial value problem

$$\begin{aligned}y'(t) &= f(t, y(t)) \\ y(t_0) &= y_0\end{aligned}$$

Here $f(t, y)$ is a given known function, t_0 is a given initial time and y_0 is a given initial value for y . The unknown in the problem is the function $y(t)$.

Two obvious considerations in deciding whether or not a given algorithm is of any practical value are

1. the amount of computational effort required to execute the algorithm and
2. the accuracy that this computational effort yields.

For algorithms like our simple ODE solvers, the bulk of the computational effort usually goes into evaluating the function¹ Typically, evaluating a complicated function will take a great many arithmetic operations, while the actual ODE solver method (as per, for example, (D.1.3)) takes only an additional handful of operations. So the great bulk of computational time goes into evaluating f and we want to do it as few times as possible. $f(t, y)$. Euler's method uses one evaluation of $f(t, y)$ for each step; the improved Euler's method uses two evaluations of f per step; the Runge-Kutta algorithm uses four evaluations of f per step. So Runge-Kutta costs four times as much work per step as does Euler. But this fact is extremely deceptive because, as we shall see, you typically get the same accuracy with a few steps of Runge-Kutta as you do with hundreds of steps of Euler.

To get a first impression of the error behaviour of these methods, we apply them to a problem that we know the answer to. The solution to the first order constant coefficient linear initial value problem

$$\begin{aligned}y'(t) &= y - 2t \\ y(0) &= 3\end{aligned}$$

is

$$y(t) = 2 + 2t + e^t$$

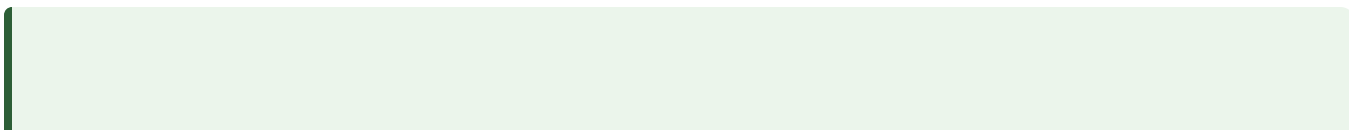
In particular, the exact value of $y(1)$, to ten decimal places, is $4 + e = 6.7182818285$. The following table lists the error in the approximate value for this number generated by our three methods applied with three different step sizes. It also lists the number of evaluations of f required to compute the approximation.

steps	Euler		Improved Euler		Runge Kutta	
	error	#evals	error	#evals	error	#evals
5	2.3×10^{-1}	5	1.6×10^{-2}	10	3.1×10^{-5}	20
50	2.7×10^{-2}	50	1.8×10^{-4}	100	3.6×10^{-9}	200
500	2.7×10^{-3}	500	1.8×10^{-6}	1000	3.6×10^{-13}	2000

Observe

- Using 20 evaluations of f worth of Runge-Kutta gives an error 90 times smaller than 500 evaluations of f worth of Euler.
- With Euler's method, decreasing the step size by a factor of ten appears to reduce the error by about a factor of ten.
- With improved Euler, decreasing the step size by a factor of ten appears to reduce the error by about a factor of one hundred.
- With Runge-Kutta, decreasing the step size by a factor of ten appears to reduce the error by about a factor of about 10^4 .

Use $A_E(h)$, $A_{IE}(h)$ and $A_{RK}(h)$ to denote the approximate value of $y(1)$ given by Euler, improved Euler and Runge-Kutta, respectively, with step size h . It looks like



Equation D.2.1

$$A_E(h) \approx y(1) + K_E h$$

$$A_{IE}(h) \approx y(1) + K_{IE} h^2$$

$$A_{RK}(h) \approx y(1) + K_{RK} h^4$$

with some constants K_E , K_{IE} and K_{RK} .

To test these conjectures further, we apply our three methods with about ten different step sizes of the form $\frac{1}{n} = \frac{1}{2^m}$ with m integer. Below are three graphs, one for each method. Each contains a plot of $Y = \log_2 e_n$, the (base 2) logarithm of the error for step size $\frac{1}{n}$, against the logarithm (of base 2) of n . The logarithm of base 2 is used because $\log_2 n = \log_2 2^m = m$ — nice and simple.

Here is why it is a good reason to plot $Y = \log_2 e_n$ against $x = \log_2 n$. If, for some algorithm, there are (unknown) constants K and k such that

$$\text{approx value of } y(1) \text{ with step size } h = y(1) + Kh^k$$

then the error with step size $\frac{1}{n}$ is $e_n = K \frac{1}{n^k}$ and obeys

$$\log_2 e_n = \log_2 K - k \log_2 n$$

The graph of $Y = \log_2 e_n$ against $x = \log_2 n$ is the straight line $Y = -kx + \log_2 K$ of slope $-k$ and y intercept $\log_2 K$.

Remark D.2.2

This procedure can still be used even if we do not know the exact value of $y(1)$. Suppose, more generally, that we have some algorithm that generates approximate values for some (unknown) exact value A . Call A_h the approximate value with step size h . Suppose that

$$A_h = A + Kh^k$$

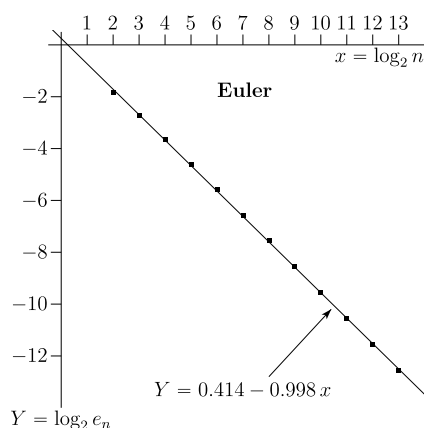
with K and k constant (but also unknown). Then plotting

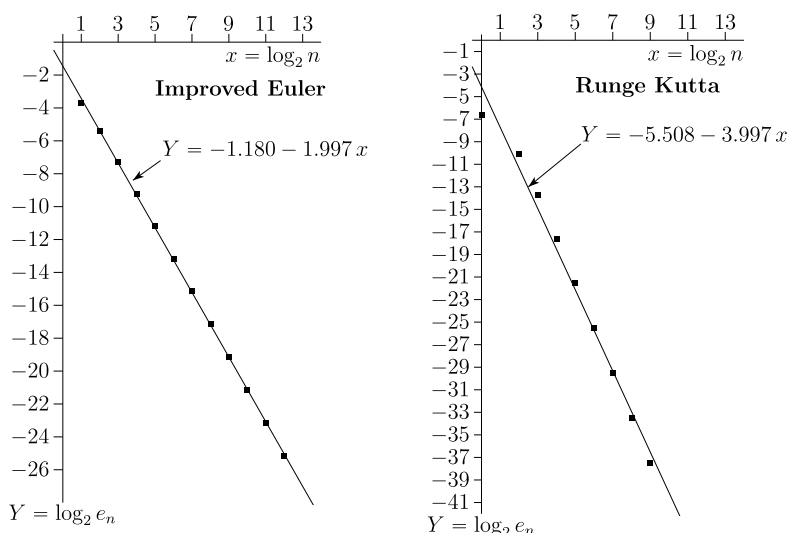
$$y = \log(A_h - A_{h/2}) = \log\left(Kh^k - K\left(\frac{h}{2}\right)^k\right) = \log\left(K - \frac{K}{2^k}\right) + k \log h$$

against $x = \log h$ gives the straight line $y = mx + b$ with slope $m = k$ and y intercept $b = \log\left(K - \frac{K}{2^k}\right)$. So we can

- read off k from the slope of the line and then
- compute $K = e^b \left(1 - \frac{1}{2^k}\right)^{-1}$ from the y intercept b and then
- compute $A = A_h - Kh^k$.

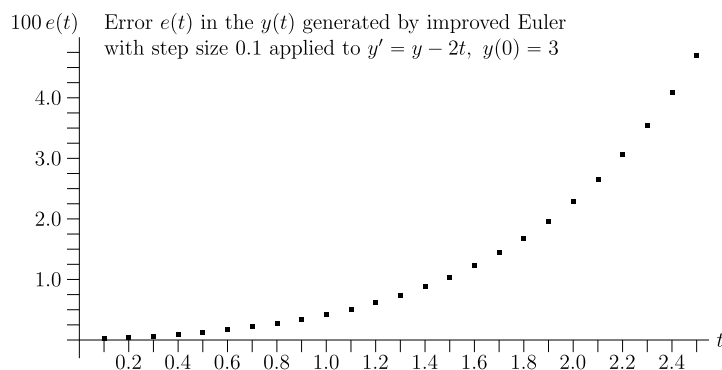
Here are the three graphs — one each for the Euler method, the improved Euler method and the Runge-Kutta method.



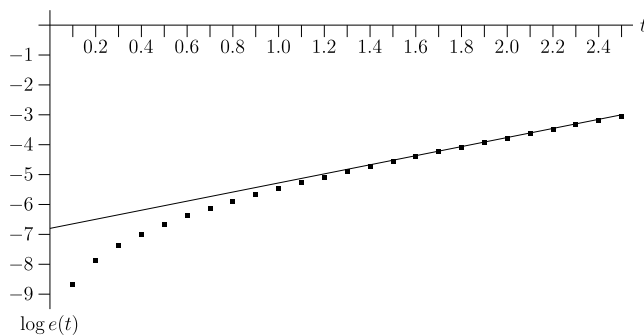


Each graph contains about a dozen data points, $(x, Y) = (\log_2 n, \log_2 e_n)$, and also contains a straight line, chosen by linear regression, to best fit the data. The method of linear regression for finding the straight line which best fits a given set of data points is covered in Example 2.9.11 of the CLP-3 text. The three straight lines have slopes -0.998 for Euler, -1.997 for improved Euler and -3.997 for Runge Kutta. Reviewing (E1), it sure looks like $k = 1$ for Euler, $k = 2$ for improved Euler and $k = 4$ for Runge-Kutta (at least if k is integer).

So far we have only looked at the error in the approximate value of $y(t_f)$ as a function of the step size h with t_f held fixed. The graph below illustrates how the error behaves as a function of t , with h held fixed. That is, we hold the step size fixed and look at the error as a function of the distance, t , from the initial point.



From the graph, it appears that the error grows exponentially with t . But it is not so easy to visually distinguish exponential curves from other upward curving curves. On the other hand, it is pretty easy to visually distinguish straight lines from other curves, and taking a logarithm converts the exponential curve $y = e^{kx}$ into the straight line $Y = \log y = kx$. Here is a graph of the logarithm, $\log e(t)$, of the error at time t , $e(t)$, against t . We have added a straight line as an aide to your eye.



It looks like the log of the error grows very quickly initially, but then settles into a straight line. Hence it really does look like, at least in this example, except at the very beginning, the error $e(t)$ grows exponentially with t .

The above numerical experiments have given a little intuition about the error behaviour of the Euler, improved Euler and Runge-Kutta methods. It's time to try and understand what is going on more rigorously.

Local Truncation Error for Euler's Method

We now try to develop some understanding as to why we got the above experimental results. We start with the error generated by a single step of Euler's method.

Definition D.2.3 Local truncation error

The (signed) error generated by a single step of Euler's method, under the assumptions that we start the step with the exact solution and that there is no roundoff error, is called the **local truncation error** for Euler's method. That is, if $\phi(t)$ obeys $\phi'(t) = f(t, \phi(t))$ and $\phi(t_n) = y_n$, and if $y_{n+1} = y_n + hf(t_n, y_n)$, then the local truncation error for Euler's method is

$$\phi(t_{n+1}) - y_{n+1}$$

That is, it is difference between the exact value, $\phi(t_{n+1})$, and the approximate value generated by a single Euler method step, y_{n+1} , ignoring any numerical issues caused by storing numbers in a computer.

Denote by $\phi(t)$ the exact solution to the initial value problem

$$y'(t) = f(t, y) \quad y(t_n) = y_n$$

That is, $\phi(t)$ obeys

$$\phi'(t) = f(t, \phi(t)) \quad \phi(t_n) = y_n$$

for all t . Now execute one more step of Euler's method with step size h :

$$y_{n+1} = y_n + hf(t_n, y_n)$$

Because we are assuming that $y_n = \phi(t_n)$

$$y_{n+1} = \phi(t_n) + hf(t_n, \phi(t_n))$$

Because $\phi(t)$ is the exact solution, $\phi'(t_n) = f(t_n, \phi(t_n)) = f(t_n, y_n)$ and

$$y_{n+1} = \phi(t_n) + h\phi'(t_n)$$

The local truncation error in y_{n+1} is, by definition, $\phi(t_{n+1}) - y_{n+1}$.

Taylor expanding (see (3.4.10) in the CLP-1 text) $\phi(t_{n+1}) = \phi(t_n + h)$ about t_n

$$\phi(t_n + h) = \phi(t_n) + \phi'(t_n)h + \frac{1}{2}\phi''(t_n)h^2 + \frac{1}{3!}\phi'''(t_n)h^3 + \dots$$

so that

$$\begin{aligned} \phi(t_{n+1}) - y_{n+1} &= [\phi(t_n) + \phi'(t_n)h + \frac{1}{2}\phi''(t_n)h^2 + \frac{1}{3!}\phi'''(t_n)h^3 + \dots] - [\phi(t_n) + h\phi'(t_n)] \\ &= \frac{1}{2}\phi''(t_n)h^2 + \frac{1}{3!}\phi'''(t_n)h^3 + \dots \end{aligned} \tag{E2}$$

Notice that the constant and h^1 terms have cancelled out. So the first term that appears is proportional to h^2 . Since h is typically a very small number, the h^3, h^4, \dots terms will usually be much smaller than the h^2 term.

We conclude that the local truncation error for Euler's method is h^2 times some unknown constant (we usually don't know the value of $\frac{1}{2}\phi''(t_n)$ because we don't usually know the solution $\phi(t)$ of the differential equation) plus smaller terms that are proportional to h^r with $r \geq 3$. This conclusion is typically written

Equation D.2.4

$$\text{Local truncation error for Euler's method} = Kh^2 + O(h^3)$$

The symbol $O(h^3)$ is used to designate any function that, for small h , is bounded by a constant times h^3 . So, if h is very small, $O(h^3)$ will be a lot smaller than h^2 .

To get from an initial time $t = t_0$ to a final time $t = t_f$ using steps of size h requires $(t_f - t_0)/h$ steps. If each step were to introduce an error³ For simplicity, we are assuming that K takes the same value in every step. If, instead, there is a different K in each of the $n = (t_f - t_0)/h$ steps, the final error would be $K_1h^2 + K_2h^2 + \dots + K_nh^2 + nO(h^3) = \bar{K}nh^2 + nO(h^3) = \bar{K}(t_f - t_0)h + O(h^2)$, where \bar{K} is the average of K_1, K_2, \dots, K_n . $Kh^2 + O(h^3)$, then the final error in the approximate value of $y(t_f)$ would be

$$\frac{t_f - t_0}{h} [Kh^2 + O(h^3)] = K(t_f - t_0)h + O(h^2)$$

This very rough estimate is consistent with the experimental data for the dependence of error on step size with t_f held fixed, shown on the first graph after Remark D.2.2. But it is not consistent with the experimental time dependence data above, which shows the error growing exponentially, rather than linearly, in $t_f - t_0$.

We can get some rough understanding of this exponential growth as follows. The general solution to $y' = y - 2t$ is $y = 2 + 2t + c_0e^t$. The arbitrary constant, c_0 , is to be determined by initial conditions. When $y(0) = 3$, $c_0 = 1$. At the end of step 1, we have computed an approximation y_1 to $y(h)$. This y_1 is not exactly $y(h) = 2 + 2h + e^h$. Instead, it is a number that differs from $2 + 2h + e^h$ by $O(h^2)$. We choose to write the number $y_1 = 2 + 2h + e^h + O(h^2)$ as $2 + 2h + (1 + \epsilon)e^h$ with $\epsilon = e^{-h}O(h^2)$ of order of magnitude h^2 . That is, we choose to write

$$y_1 = 2 + 2t + c_0e^t \Big|_{t=h} \quad \text{with } c_0 = 1 + \epsilon$$

If we were to make no further errors we would end up with the solution to

$$y' = y - 2t, \quad y(h) = 2 + 2h + (1 + \epsilon)e^h$$

which is⁴

$$\begin{aligned} y(t) &= 2 + 2t + (1 + \epsilon)e^t = 2 + 2t + e^t + \epsilon e^t \\ &= \phi(t) + \epsilon e^t \end{aligned}$$

So, once as error has been introduced, the natural time evolution of the solutions to this differential equation cause the error to grow exponentially. Other differential equations with other time evolution characteristics will exhibit different t_f dependence of errors⁵ For example, if the solution is polynomial, then we might expect (by a similar argument) that the error also grows polynomially in t_f . In the next section, we show that, for many differential equations, errors grow at worst exponentially with t_f .

Global Truncation Error for Euler's Method

Suppose once again that we are applying Euler's method with step size h to the initial value problem

$$\begin{aligned} y'(t) &= f(t, y) \\ y(0) &= y_0 \end{aligned}$$

Denote by $\phi(t)$ the exact solution to the initial value problem and by y_n the approximation to $\phi(t_n)$, $t_n = t_0 + nh$, given by n steps of Euler's method (applied without roundoff error).

Definition D.2.5 Global truncation error

The (signed) error in y_n is $\phi(t_n) - y_n$ and is called the **global truncation error** at time t_n .

The word "truncation" is supposed to signify that this error is due solely to Euler's method and does not include any effects of roundoff error that might be introduced by our not writing down an infinite number of decimal digits for each number that we compute along the way. We now derive a bound on the global truncation error.

Define

$$\varepsilon_n = \phi(t_n) - y_n$$

The first half of the derivation is to find a bound on ε_{n+1} in terms of ε_n .

$$\begin{aligned} \varepsilon_{n+1} &= \phi(t_{n+1}) - y_{n+1} \\ &= \phi(t_{n+1}) - y_n - hf(t_n, y_n) \\ &= [\phi(t_n) - y_n] + h[f(t_n, \phi(t_n)) - f(t_n, y_n)] \\ &\quad + [\phi(t_{n+1}) - \phi(t_n) - hf(t_n, \phi(t_n))] \end{aligned} \tag{E3}$$

where we have massaged the expression into three manageable pieces.

- The first $[\dots]$ is exactly ε_n .
- The third $[\dots]$ is exactly the local truncation error. Assuming that $|\phi''(t)| \leq A$ for all t of interest ⁶We are assuming that the derivative $\phi'(t)$ doesn't change too rapidly. This will be the case if $f(t, y)$ is a reasonably smooth function., we can bound the third $[\dots]$ by

$$|\phi(t_{n+1}) - \phi(t_n) - hf(t_n, \phi(t_n))| \leq \frac{1}{2}Ah^2$$

This bound follows quickly from the Taylor expansion with remainder ((3.4.32) in the CLP-1 text),

$$\begin{aligned} \phi(t_{n+1}) &= \phi(t_n) + \phi'(t_n)h + \frac{1}{2}\phi''(\tilde{t})h^2 \\ &= \phi(t_n) + hf(t_n, \phi(t_n)) + \frac{1}{2}\phi''(\tilde{t})h^2 \end{aligned}$$

for some $t_n < \tilde{t} < t_{n+1}$.

- Finally, by the mean value theorem, the magnitude of the second $[\dots]$ is h times

$$\begin{aligned} |f(t_n, \phi(t_n)) - f(t_n, y_n)| &= F_{t_n}(\phi(t_n)) - F_{t_n}(y_n) \quad \text{where } F_{t_n}(y) = f(t_n, y) \\ &= |F'_{t_n}(\tilde{y})| |\phi(t_n) - y_n| \\ &\quad \text{for some } \tilde{y} \text{ between } y_n \text{ and } \phi(t_n) \\ &= |F'_{t_n}(\tilde{y})| |\varepsilon_n| \\ &\leq B|\varepsilon_n| \end{aligned}$$

assuming that $|F'_t(y)| \leq B$ for all t and y of interest ⁷Again, this will be the case if $f(t, y)$ is a reasonably smooth function..

Substituting into (E3) gives

$$|\varepsilon_{n+1}| \leq |\varepsilon_n| + Bh|\varepsilon_n| + \frac{1}{2}Ah^2 = (1 + Bh)|\varepsilon_n| + \frac{1}{2}Ah^2$$

Hence the (bound on the) magnitude of the total error, $|\varepsilon_{n+1}|$, consists of two parts. One part is the magnitude of the local truncation error, which is no more than $\frac{1}{2}Ah^2$ and which is present even if we start the step with no error at all, i.e. with $\varepsilon_n = 0$. The other part is due to the combined error from all previous steps. This is the ε_n term. At the beginning of step number $n + 1$, the combined error has magnitude $|\varepsilon_n|$. During the step, this error gets magnified by no more than a factor of $1 + Bh$.

The second half of the derivation is to repeatedly apply (E4-n) with $n = 0, 1, 2, \dots$. By definition $\phi(t_0) = y_0$ so that $\varepsilon_0 = 0$, so

$$\begin{aligned} \text{(E4-0)} &\implies |\varepsilon_1| \leq (1 + Bh)|\varepsilon_0| + \frac{A}{2}h^2 = \frac{A}{2}h^2 \\ \text{(E4-1)} &\implies |\varepsilon_2| \leq (1 + Bh)|\varepsilon_1| + \frac{A}{2}h^2 = (1 + Bh)\frac{A}{2}h^2 + \frac{A}{2}h^2 \\ \text{(E4-2)} &\implies |\varepsilon_3| \leq (1 + Bh)|\varepsilon_2| + \frac{A}{2}h^2 = (1 + Bh)^2\frac{A}{2}h^2 + (1 + Bh)\frac{A}{2}h^2 + \frac{A}{2}h^2 \end{aligned}$$

Continuing in this way

$$|\varepsilon_n| \leq (1 + Bh)^{n-1}\frac{A}{2}h^2 + \dots + (1 + Bh)\frac{A}{2}h^2 + \frac{A}{2}h^2 = \sum_{m=0}^{n-1} (1 + Bh)^m \frac{A}{2}h^2$$

This is the beginning of a geometric series, and we can sum it up by using $\sum_{m=0}^{n-1} ar^m = \frac{r^n - 1}{r - 1}a$ (which is Theorem 1.1.6(a)) with

$a = \frac{A}{2}h^2$ and $r = 1 + Bh$ gives

$$|\varepsilon_n| \leq \frac{(1+Bh)^n - 1}{(1+Bh) - 1} \frac{A}{2} h^2 = \frac{A}{2B} [(1+Bh)^n - 1] h$$

We are interested in how this behaves as $t_n - t_0$ increases and/or h decreases. Now $n = \frac{t_n - t_0}{h}$ so that $(1+Bh)^n = (1+Bh)^{(t_n - t_0)/h}$. When h is small, the behaviour of $(1+Bh)^{(t_n - t_0)/h}$ is not so obvious. So we'll use a little trickery to make it easier to understand. Setting $x = Bh$ in

$$x \geq 0 \implies 1 + x \leq 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots = e^x$$

(the exponential series $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots$ was derived in Example 3.6.5. gives⁸ When $x = Bh$ is large, it is not wise to bound the linear $1 + x$ by the much larger exponential e^x . However when x is small, $1 + x$ and e^x are almost the same. $1 + Bh \leq e^{Bh}$. Hence $(1+Bh)^n \leq e^{Bhn} = e^{B(t_n - t_0)}$, since $t_n = t_0 + nh$, and we arrive at the conclusion

Equation D.2.6

$$|\varepsilon_n| \leq \frac{A}{2B} [e^{B(t_n - t_0)} - 1] h$$

This is of the form $K(t_f)h^k$ with $k = 1$ and the coefficient $K(t_f)$ growing exponentially with $t_f - t_0$. If we keep h fixed and increase t_n we see exponential growth, but if we fix t_n and decrease h we see the error decrease linearly. This is just what our experimental data suggested.

This page titled D.2: Simple ODE Solvers — Error Behaviour is shared under a [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

D.3: Variable Step Size Methods

- We now introduce a family of procedures that decide by themselves what step size to use. In all of these procedures the user specifies an acceptable error rate and the procedure attempts to adjust the step size so that each step introduces error at no more than that rate. That way the procedure uses a small step size when it is hard to get an accurate approximation, and a large step size when it is easy to get a good approximation.

Suppose that we wish to generate an approximation to the initial value problem

$$y' = f(t, y), \quad y(t_0) = y_0$$

for some range of t 's and we want the error introduced per unit increase¹ We know that the error will get larger the further we go in t . So it makes sense to try to limit the error per unit increase in t . of t to be no more than about some small fixed number ε . This means that if $y_n \approx y(t_0 + nh)$ and $y_{n+1} \approx y(t_0 + (n+1)h)$, then we want the local truncation error in the step from y_n to y_{n+1} to be no more than about εh . Suppose further that we have already produced the approximate solution as far as t_n . The rough strategy is as follows. We do the step from t_n to $t_n + h$ twice using two different algorithms, giving two different approximations to $y(t_{n+1})$, that we call $A_{1,n+1}$ and $A_{2,n+1}$. The two algorithms are chosen so that

1. we can use $A_{1,n+1} - A_{2,n+1}$ to compute an approximate local truncation error and
2. for efficiency, the two algorithms use almost the same evaluations of f . Remember that evaluating the function f is typically the most time-consuming part of our computation.

In the event that the local truncation error, divided by h , (i.e. the error per unit increase of t) is smaller than ε , we set $t_{n+1} = t_n + h$, accept $A_{2,n+1}$ as the approximate value²Better still, accept $A_{2,n+1}$ minus the computed approximate error in $A_{2,n+1}$ as the approximate value for $y(t_{n+1})$, for $y(t_{n+1})$, and move on to the next step. Otherwise we pick, using what we have learned from $A_{1,n+1} - A_{2,n+1}$, a new trial step size h and start over again at t_n .

Now for the details. We start with a very simple procedure. We will later soup it up to get a much more efficient procedure.

Euler and Euler-2step (preliminary version)

Denote by $\phi(t)$ the exact solution to $y' = f(t, y)$ that satisfies the initial condition $\phi(t_0) = y_0$. If we apply one step of Euler with step size h , giving

$$A_{1,n+1} = y_n + hf(t_n, y_n)$$

we know, from (D.2.4), that

$$A_{1,n+1} = \phi(t_n + h) + Kh^2 + O(h^3)$$

The problem, of course, is that we don't know what the error is, even approximately, because we don't know what the constant K is. But we can estimate K simply by redoing the step from t_n to $t_n + h$ using a judiciously chosen second algorithm. There are a number of different second algorithms that will work. We call the simple algorithm that we use in this subsection Euler-2step³This name is begging for a dance related footnote and we invite the reader to supply their own.. One step of Euler-2step with step size h just consists of doing two steps of Euler with step size $h/2$:

$$A_{2,n+1} = y_n + \frac{h}{2}f(t_n, y_n) + \frac{h}{2}f(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n))$$

Here, the first half-step took us from y_n to $y_{\text{mid}} = y_n + \frac{h}{2}f(t_n, y_n)$ and the second half-step took us from y_{mid} to $y_{\text{mid}} + \frac{h}{2}f(t_n + \frac{h}{2}, y_{\text{mid}})$. The local truncation error introduced in the first half-step is $K(h/2)^2 + O(h^3)$. That for the second half-step is $K(h/2)^2 + O(h^3)$ with the same⁴Because the two half-steps start at values of t only $h/2$ apart, and we are thinking of h as being very small, it should not be surprising that we can use the same value of K in both. In case you don't believe us, we have included a derivation of the local truncation error for Euler-2step later in this appendix. K , though with a different $O(h^3)$. All together

$$\begin{aligned} A_{2,n+1} &= \phi(t_n + h) + [K(\frac{h}{2})^2 + O(h^3)] + [K(\frac{h}{2})^2 + O(h^3)] \\ &= \phi(t_n + h) + \text{half}Kh^2 + O(h^3) \end{aligned}$$

The difference is⁵

$$\begin{aligned} A_{1,n+1} - A_{2,n+1} &= [\phi(t_n + h) + Kh^2 + O(h^3)] - [\phi(t_n + h) - \frac{1}{2}Kh^2 - O(h^3)] \\ &= \frac{1}{2}Kh^2 + O(h^3) \end{aligned}$$

So if we do one step of both Euler and Euler-2step, we can estimate

$$\frac{1}{2}Kh^2 = A_{1,n+1} - A_{2,n+1} + O(h^3)$$

We now know that in the step just completed Euler-2step introduced an error of about $\frac{1}{2}Kh^2 \approx A_{1,n+1} - A_{2,n+1}$. That is, the current error rate is about $r = \frac{|A_{1,n+1} - A_{2,n+1}|}{h} \approx \frac{1}{2}|K|h$ per unit increase of t .

- If this $r = \frac{|A_{1,n+1} - A_{2,n+1}|}{h} > \varepsilon$, we reject⁶ The measured error rate, r , is bigger than the desired error rate ε . That means that it is harder to get the accuracy we want than we thought. So we have to take smaller steps. $A_{2,n+1}$ and repeat the current step with a new trial step size chosen so that $\frac{1}{2}|K|(\text{new } h) < \varepsilon$, i.e. $\frac{r}{h}(\text{new } h) < \varepsilon$. To give ourselves a small safety margin, we could use⁷ We don't want to make the new h too close to $\frac{\varepsilon}{r}h$ since we are only estimating things and we might end up with an error rate bigger than ε . On the other hand, we don't want to make the new h too small because that means too much work — so we choose it to be just a little smaller than $\frac{\varepsilon}{r}h$. . . say $0.9\frac{\varepsilon}{r}h$.

$$\text{new } h = 0.9 \frac{\varepsilon}{r} h$$

- If $r = \frac{|A_{1,n+1} - A_{2,n+1}|}{h} < \varepsilon$ we can accept⁸ The measured error rate, r , is smaller than the desired error rate ε . That means that it is easier to get the accuracy we want than we thought. So we can make the next step larger. $A_{2,n+1}$ as an approximate value for $y(t_{n+1})$, with $t_{n+1} = t_n + h$, and move on to the next step, starting with the new trial step size⁹ Note that in this case $\frac{\varepsilon}{r} > 1$. So the new h can be bigger than the last h .

$$\text{new } h = 0.9 \frac{\varepsilon}{r} h$$

That is our preliminary version of the Euler/Euler-2step variable step size method. We call it the preliminary version, because we will shortly tweak it to get a much more efficient procedure.

✓ Example D.3.1

As a concrete example, suppose that our problem is

$$y(0) = e^{-2}, \quad y' = 8(1 - 2t)y, \quad \varepsilon = 0.1$$

and that we have gotten as far as

$$t_n = 0.33, \quad y_n = 0.75, \quad \text{trial } h = 0.094$$

Then, using $E = |A_{1,n+1} - A_{2,n+1}|$ to denote the magnitude of the estimated local truncation error in $A_{2,n+1}$ and r the corresponding error rate

$$\begin{aligned} f(t_n, y_n) &= 8(1 - 2 \times 0.33)0.75 = 2.04 \\ A_{1,n+1} &= y_n + hf(t_n, y_n) = 0.75 + 0.094 \times 2.04 = 0.942 \\ y_{\text{mid}} &= y_n + \frac{h}{2}f(t_n, y_n) = 0.75 + \frac{0.094}{2} \times 2.04 = 0.846 \\ f(t_n + \frac{h}{2}, y_{\text{mid}}) &= 8 \left[1 - 2 \left(0.33 + \frac{0.094}{2} \right) \right] 0.846 = 1.66 \\ A_{2,n+1} &= y_{\text{mid}} + \frac{h}{2}f(t_n + \frac{h}{2}, y_{\text{mid}}) = 0.846 + \frac{0.094}{2} 1.66 = 0.924 \\ E &= |A_{1,n+1} - A_{2,n+1}| = |0.942 - 0.924| = 0.018 \\ r &= \frac{|E|}{h} = \frac{0.018}{0.094} = 0.19 \end{aligned}$$

Since $r = 0.19 > \varepsilon = 0.1$, the current step size is unacceptable and we have to recompute with the new step size

$$\text{new } h = 0.9 \frac{\varepsilon}{r} (\text{old } h) = 0.9 \frac{0.1}{0.19} 0.094 = 0.045$$

to give

$$\begin{aligned}
 f(t_n, y_n) &= 8(1 - 2 \times 0.33)0.75 = 2.04 \\
 A_{1,n+1} &= y_n + hf(t_n, y_n) = 0.75 + 0.045 \times 2.04 = 0.842 \\
 y_{\text{mid}} &= y_n + \frac{h}{2} f(t_n, y_n) = 0.75 + \frac{0.045}{2} \times 2.04 = 0.796 \\
 f(t_n + \frac{h}{2}, y_{\text{mid}}) &= 8 \left[1 - 2 \left(0.33 + \frac{0.045}{2} \right) \right] 0.796 = 1.88 \\
 A_{2,n+1} &= y_{\text{mid}} + \frac{h}{2} f(t_n + \frac{h}{2}, y_{\text{mid}}) = 0.796 + \frac{0.045}{2} 1.88 = 0.838 \\
 E &= A_{1,n+1} - A_{2,n+1} = 0.842 - 0.838 = 0.004 \\
 r &= \frac{|E|}{h} = \frac{0.004}{0.045} = 0.09
 \end{aligned}$$

This time $r = 0.09 < \varepsilon = 0.1$, is acceptable so we set $t_{n+1} = 0.33 + 0.045 = 0.375$ and

$$y_{n+1} = A_{2,n+1} = 0.838$$

The initial trial step size from t_{n+1} to t_{n+2} is

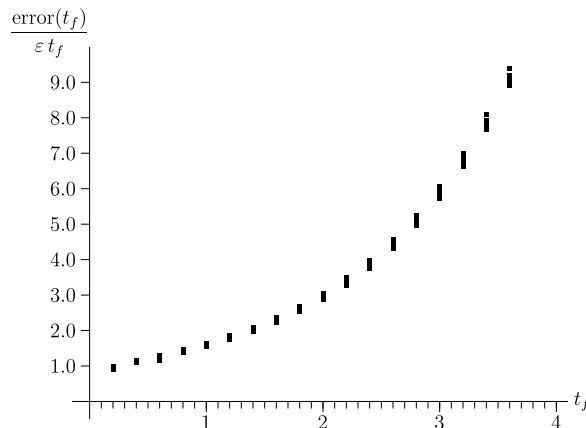
$$\text{new } h = 0.9 \frac{\varepsilon}{r} (\text{old } h) = 0.9 \frac{0.1}{0.09} \cdot 0.045 = .045$$

By a fluke, it has turned out that the new h is the same as the old h (to three decimal places). If r had been significantly smaller than ε , then the new h would have been significantly bigger than the old h - indicating that it is (relatively) easy to estimate things in this region, making a larger step size sufficient.

As we said above, we will shortly upgrade the above variable step size method, that we are calling the preliminary version of the Euler/Euler-2step method, to get a much more efficient procedure. Before we do so, let's pause to investigate a little how well our preliminary procedure does at controlling the rate of error production.

We have been referring, loosely, to ε as the desired rate for introduction of error, by our variable step size method, as t advances. If the rate of increase of error were exactly ε , then at final time t_f the accumulated error would be exactly $\varepsilon(t_f - t_0)$. But our algorithm actually chooses the step size h for each step so that the estimated local truncation error in $A_{2,n+1}$ for that step is about εh . We have seen that, once some local truncation error has been introduced, its contribution to the global truncation error can grow exponentially with t_f .

Here are the results of a numerical experiment that illustrate this effect. In this experiment, the above preliminary Euler/Euler-2step method is applied to the initial value problem $y' = t - 2y$, $y(0) = 3$ for $\varepsilon = \frac{1}{16}, \frac{1}{32}, \dots$ (ten different values) and for $t_f = 0.2, 0.4, \dots, 3.8$. Here is a plot of the resulting $\frac{\text{actual error at } t=t_f}{\varepsilon t_f}$ against t_f .



If the rate of introduction of error were exactly ε , we would have $\frac{\text{actual error at } t=t_f}{\varepsilon t_f} = 1$. There is a small square on the graph for each different pair ε, t_f . So for each value of t_f there are ten (possibly overlapping) squares on the line $x = t_f$. This numerical experiment suggests that $\frac{\text{actual error at } t=t_f}{\varepsilon t_f}$ is relatively independent of ε and starts, when t_f is small, at about one, as we want, but grows (perhaps exponentially) with t_f .

Euler and Euler-2step (final version)

We are now ready to use a sneaky bit of arithmetic to supercharge our Euler/Euler-2step method. As in our development of the preliminary version of the method, denote by $\phi(t)$ the exact solution to $y' = f(t, y)$ that satisfies the initial condition $\phi(t_n) = y_n$. We have seen, at the beginning of §D.3.1, that applying one step of Euler with step size h , gives

$$\begin{aligned} A_{1,n+1} &= y_n + hf(t_n, y_n) \\ &= \phi(t_n + h) + Kh^2 + O(h^3) \end{aligned} \tag{E5}$$

and applying one step of Euler-2step with step size h (i.e. applying two steps of Euler with step size $h/2$) gives

$$\begin{aligned} A_{2,n+1} &= y_n + \frac{h}{2}f(t_n, y_n) + \frac{h}{2}f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)\right) \\ &= \phi(t_n + h) + \frac{1}{2}Kh^2 + O(h^3) \end{aligned} \tag{E6}$$

because the local truncation error introduced in the first half-step was $K(h/2)^2 + O(h^3)$ and that introduced in the second half-step was $K(h/2)^2 + O(h^3)$. Now here is the sneaky bit. Equations (E5) and (E6) are very similar and we can eliminate all Kh^2 's by subtracting (E5) from 2 times (E6). This gives

$$2(\text{E6}) - (\text{E5}): \quad 2A_{2,n+1} - A_{1,n+1} = \phi(t_n + h) + O(h^3)$$

(no more h^2 term!) or

$$\phi(t_n + h) = 2A_{2,n+1} - A_{1,n+1} + O(h^3)$$

which tells us that choosing

$$y_{n+1} = 2A_{2,n+1} - A_{1,n+1}$$

would give a local truncation error of order h^3 , rather than the order h^2 of the preliminary Euler/Euler-2step method. To convert the preliminary version to the final version, we just replace $y_{n+1} = A_{2,n+1}$ by $y_{n+1} = 2A_{2,n+1} - A_{1,n+1}$:

Equation D.3.2 Euler/Euler-2step Method

Given $\varepsilon > 0$, t_n , y_n and the current step size h

- compute

$$\begin{aligned} A_{1,n+1} &= y_n + hf(t_n, y_n) \\ A_{2,n+1} &= y_n + \frac{h}{2}f(t_n, y_n) + \frac{h}{2}f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)\right) \\ r &= \frac{|A_{1,n+1} - A_{2,n+1}|}{h} \end{aligned}$$

- If $r > \varepsilon$, repeat the first bullet but with the new step size

$$(\text{new } h) = 0.9 \frac{\varepsilon}{r} (\text{old } h)$$

- If $r < \varepsilon$ set

$$\begin{aligned} t_{n+1} &= t_n + h \\ y_{n+1} &= 2A_{2,n+1} - A_{1,n+1} \quad \text{and the new trial step size} \\ (\text{new } h) &= 0.9 \frac{\varepsilon}{r} (\text{old } h) \end{aligned}$$

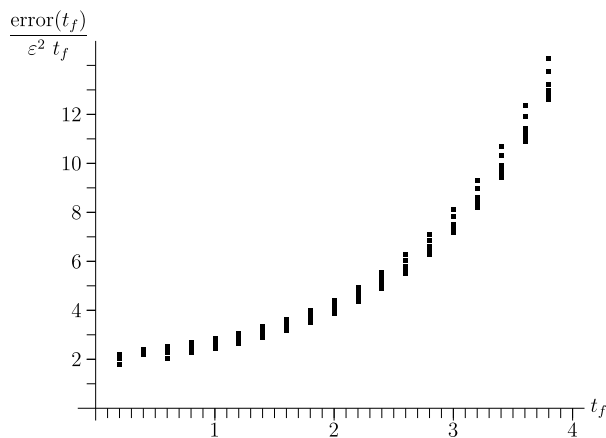
and move on to the next step. Note that since $r < \varepsilon$, $\frac{\varepsilon}{r}h > h$ which indicates that the new h can be larger than the old h . We include the 0.9 to be careful not to make the error of the next step too big.

Let's think a bit about how our final Euler/Euler-2step method should perform.

- The step size here, as in the preliminary version, is chosen so that the local truncation error in $A_{2,n+1}$ per unit increase of t , namely $r = \frac{|A_{1,n+1} - A_{2,n+1}|}{h} \approx \frac{Kh^2/2}{h} = \frac{K}{2}h$, is approximately ε . So h is roughly proportional to ε .

- On the other hand, (E7) shows that, in the full method, local truncation error is being added to y_{n+1} at a rate of $\frac{O(h^3)}{h} = O(h^2)$ per unit increase in t .
- So one would expect that local truncation increases the error at a rate proportional to ε^2 per unit increase in t .
- If the rate of increase of error were exactly a constant time ε^2 , then the error accumulated between the initial time $t = 0$ and the final time $t = t_f$ would be exactly a constant times $\varepsilon^2 t_f$.
- However we have seen that, once some local truncation error has been introduced, its contribution to the global error can grow exponentially with t_f . So we would expect that, under the full Euler/Euler-2step method, $\frac{\text{actual error at } t=t_f}{\varepsilon^2 t_f}$ to be more or less independent of ε , but still growing exponentially in t_f .

Here are the results of a numerical experiment that illustrate this. In this experiment, the above final Euler/Euler-2step method, (D.3.2), is applied to the initial value problem $y' = t - 2y$, $y(0) = 3$ for $\varepsilon = \frac{1}{16}, \frac{1}{32}, \dots$ (ten different values) and for $t_f = 0.2, 0.4, \dots, 3.8$. In the following plot, there is a small square for the resulting $\frac{\text{actual error at } t=t_f}{\varepsilon^2 t_f}$ for each different pair ε, t_f .



It does indeed look like $\frac{\text{actual error at } t=t_f}{\varepsilon^2 t_f}$ is relatively independent of ε but grows (perhaps exponentially) with t_f . Note that $\frac{\text{actual error at } t=t_f}{\varepsilon^2 t_f}$ contains a factor of ε^2 in the denominator. The actual error rate $\frac{\text{actual error at } t=t_f}{t_f}$ is much smaller than is suggested by the graph.

Fehlberg's Method

Of course, in practice more efficient and more accurate methods¹⁰ There are a very large number of such methods. We will only look briefly at a couple of the simpler ones. The interested reader can find more by search enging for such keywords as “Runge-Kutta methods” and “adaptive step size”. than Euler and Euler-2step are used. Fehlberg's method¹¹E. Fehlberg, NASA Technical Report R315 (1969) and NASA Technical Report R287 (1968). uses improved Euler and a second more accurate method. Each step involves three calculations of f :

$$\begin{aligned} f_{1,n} &= f(t_n, y_n) \\ f_{2,n} &= f(t_n + h, y_n + h f_{1,n}) \\ f_{3,n} &= f\left(t_n + \frac{h}{2}, y_n + \frac{h}{4} [f_{1,n} + f_{2,n}]\right) \end{aligned}$$

Once these three evaluations have been made, the method generates two approximations for $y(t_n + h)$:

$$\begin{aligned} A_{1,n+1} &= y_n + \frac{h}{2} [f_{1,n} + f_{2,n}] \\ A_{2,n+1} &= y_n + \frac{h}{6} [f_{1,n} + f_{2,n} + 4f_{3,n}] \end{aligned}$$

Denote by $\phi(t)$ the exact solution to $y' = f(t, y)$ that satisfies the initial condition $\phi(t_n) = y_n$. Now $A_{1,n+1}$ is just the y_{n+1} produced by the improved Euler's method. The local truncation error for the improved Euler's method is of order h^3 , one power of h smaller than that for Euler's method. So

$$A_{1,n+1} = \phi(t_n + h) + Kh^3 + O(h^4)$$

and it turns out¹²The interested reader can find Fehlberg's original paper online (at NASA!) and follow the derivation. It requires careful Taylor expansions and then clever algebra to cancel the bigger error terms. that

$$A_{2,n+1} = \phi(t_n + h) + O(h^4)$$

So the error in $A_{1,n+1}$ is

$$\begin{aligned} E &= |Kh^3 + O(h^4)| = |A_{1,n+1} - \phi(t_n + h)| + O(h^4) \\ &= |A_{1,n+1} - A_{2,n+1}| + O(h^4) \end{aligned}$$

and our estimate for rate at which error is being introduced into $A_{1,n+1}$ is

$$r = \frac{|A_{1,n+1} - A_{2,n+1}|}{h} \approx |K|h^2$$

per unit increase of t .

- If $r > \varepsilon$ we redo this step with a new trial step size chosen so that $|K|(\text{new } h)^2 < \varepsilon$, i.e. $\frac{r}{h^2}(\text{new } h)^2 < \varepsilon$. With our traditional safety factor, we take

$$\text{new } h = 0.9 \sqrt{\frac{\varepsilon}{r}} h \quad (\text{the new } h \text{ is smaller})$$

- If $r \leq \varepsilon$ we set $t_{n+1} = t_n + h$ and $y_{n+1} = A_{2,n+1}$ (since $A_{2,n+1}$ should be considerably more accurate than $A_{1,n+1}$) and move on to the next step with trial step size

$$\text{new } h = 0.9 \sqrt{\frac{\varepsilon}{r}} h \quad (\text{the new } h \text{ is usually bigger})$$

The Kutta-Merson Process

The Kutta-Merson process¹³R.H. Merson, "An operational method for the study of integration processes", Proc. Symp. Data Processing, Weapons Res. Establ. Salisbury, Salisbury (1957) pp. 110–125. uses two variations of the Runge-Kutta method. Each step involves five calculations¹⁴Like the other methods described above, the coefficients $1/3, 1/6, 1/8$ etc. are chosen so as to cancel larger error terms. While determining the correct choice of coefficients is not conceptually difficult, it does take some work and is beyond the scope of this appendix. The interested reader should search-engine their way to a discussion of adaptive Runge-Kutta methods. of f :

$$\begin{aligned} k_{1,n} &= f(t_n, y_n) \\ k_{2,n} &= f\left(t_n + \frac{1}{3}h, y_n + \frac{1}{3}hk_{1,n}\right) \\ k_{3,n} &= f\left(t_n + \frac{1}{3}h, y_n + \frac{1}{6}hk_{1,n} + \frac{1}{6}hk_{2,n}\right) \\ k_{4,n} &= f\left(t_n + \frac{1}{2}h, y_n + \frac{1}{8}hk_{1,n} + \frac{3}{8}hk_{3,n}\right) \\ k_{5,n} &= f\left(t_n + h, y_n + \frac{1}{2}hk_{1,n} - \frac{3}{2}hk_{3,n} + 2hk_{4,n}\right) \end{aligned}$$

Once these five evaluations have been made, the process generates two approximations for $y(t_n + h)$:

$$\begin{aligned} A_{1,n+1} &= y_n + h \left[\frac{1}{2}k_{1,n} - \frac{3}{2}k_{3,n} + 2k_{4,n} \right] \\ A_{2,n+1} &= y_n + h \left[\frac{1}{6}k_{1,n} + \frac{2}{3}k_{4,n} + \frac{1}{6}k_{5,n} \right] \end{aligned}$$

The (signed) error in $A_{1,n+1}$ is $\frac{1}{120}h^5K + O(h^6)$ while that in $A_{2,n+1}$ is $\frac{1}{720}h^5K + O(h^6)$ with the same constant K . So $A_{1,n+1} - A_{2,n+1} = \frac{5}{720}Kh^5 + O(h^6)$ and the unknown constant K can be determined, to within an error $O(h)$, by

$$K = \frac{720}{5h^5} (A_{1,n+1} - A_{2,n+1})$$

and the approximate (signed) error in $A_{2,n+1}$ and its corresponding rate per unit increase of t are

$$E = \frac{1}{720} K h^5 = \frac{1}{5} (A_{1,n+1} - A_{2,n+1})$$

$$r = \frac{|E|}{h} = \frac{1}{720} |K| h^4 = \frac{1}{5} h |A_{1,n+1} - A_{2,n+1}|$$

- If $r > \varepsilon$ we redo this step with a new trial step size chosen so that $\frac{1}{720} |K| (\text{new } h)^4 < \varepsilon$, i.e. $\frac{r}{h^4} (\text{new } h)^4 < \varepsilon$. With our traditional safety factor, we take

$$\text{new } h = 0.9 \left(\frac{\varepsilon}{r} \right)^{1/4} h$$

- If $r \leq \varepsilon$ we set $t_{n+1} = t_n + h$ and $y_{n+1} = A_{2,n+1} - E$ (since E is our estimate of the signed error in $A_{2,n+1}$) and move on to the next step with trial step size

$$\text{new } h = 0.9 \left(\frac{\varepsilon}{r} \right)^{1/4} h$$

The Local Truncation Error for Euler-2step

In our description of Euler/Euler-2step above we simply stated the local truncation error without an explanation. In this section, we show how it may be derived. We note that very similar calculations underpin the other methods we have described.

In this section, we will be using partial derivatives and, in particular, the chain rule for functions of two variables. That material is covered in Chapter 2 of the CLP-3 text. If you are not yet comfortable with it, you can either take our word for those bits, or you can delay reading this section until you have learned a little multivariable calculus.

Recall that, by definition, the local truncation error for an algorithm is the (signed) error generated by a single step of the algorithm, under the assumptions that we start the step with the exact solution and that there is no roundoff error¹⁵ We should note that in serious big numerical computations, one really does have to take rounding errors into account because they can cause serious problems. The interested reader should search-engine their way to the story of Edward Lorenz's numerical simulations and the beginnings of chaos theory. Unfortunately we simply do not have space in this text to discuss all aspects of mathematics.. Denote by $\phi(t)$ the exact solution to

$$y'(t) = f(t, y)$$

$$y(t_n) = y_n$$

In other words, $\phi(t)$ obeys

$$\phi'(t) = f(t, \phi(t)) \quad \text{for all } t$$

$$\phi(t_n) = y_n$$

In particular $\phi'(t_n) = f(t_n, \phi(t_n)) = f(t_n, y_n)$ and, carefully using the chain rule, which is (2.4.2) in the CLP-3 text,

$$\begin{aligned} \phi''(t_n) &= \left. \frac{d}{dt} f(t, \phi(t)) \right|_{t=t_n} = \left[f_t(t, \phi(t)) + f_y(t, \phi(t)) \phi'(t) \right]_{t=t_n} \\ &= f_t(t_n, y_n) + f_y(t_n, y_n) f(t_n, y_n) \end{aligned} \tag{E9}$$

Remember that f_t is the partial derivative of f with respect to t , and that f_y is the partial derivative of f with respect to y . We'll need (E9) below.

By definition, the local truncation error for Euler is

$$E_1(h) = \phi(t_n + h) - y_n - h f(t_n, y_n)$$

while that for Euler-2step is

$$E_2(h) = \phi(t_n + h) - y_n - \frac{h}{2} f(t_n, y_n) - \frac{h}{2} f(t_n + \frac{h}{2}, y_n + \frac{h}{2} f(t_n, y_n))$$

To understand how $E_1(h)$ and $E_2(h)$ behave for small h we can use Taylor expansions ((3.4.10) in the CLP-1 text) to write them as power series in h . To be precise, we use

$$g(h) = g(0) + g'(0) h + \frac{1}{2} g''(0) h^2 + O(h^3)$$

to expand both $E_1(h)$ and $E_2(h)$ in powers of h to order h^2 . Note that, in the expression for $E_1(h)$, t_n and y_n are constants — they do not vary with h . So computing derivatives of $E_1(h)$ with respect to h is actually quite simple.

$$\begin{aligned} E_1(h) &= \phi(t_n + h) - y_n - hf(t_n, y_n) & E_1(0) &= \phi(t_n) - y_n = 0 \\ E_1'(h) &= \phi'(t_n + h) - f(t_n, y_n) & E_1'(0) &= \phi'(t_n) - f(t_n, y_n) = 0 \\ E_1''(h) &= \phi''(t_n + h) & E_1''(0) &= \phi''(t_n) \end{aligned}$$

By Taylor, the local truncation error for Euler obeys

Equation D.3.3

$$E_1(h) = \frac{1}{2}\phi''(t_n)h^2 + O(h^3) = Kh^2 + O(h^3) \quad \text{with } K = \frac{1}{2}\phi''(t_n)$$

Computing arguments of $E_2(h)$ with respect to h is a little harder, since h now appears in the arguments of the function f . As a consequence, we have to include some partial derivatives.

$$\begin{aligned} E_2(h) &= \phi(t_n + h) - y_n - \frac{h}{2}f(t_n, y_n) - \frac{h}{2}f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)\right) \\ E_2'(h) &= \phi'(t_n + h) - \frac{1}{2}f(t_n, y_n) - \frac{1}{2}f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)\right) \\ &\quad - \frac{h}{2} \underbrace{\frac{d}{dh}f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)\right)}_{\text{leave this expression as is for now}} \\ E_2''(h) &= \phi''(t_n + h) - 2 \times \frac{1}{2} \underbrace{\frac{d}{dh}f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)\right)}_{\text{leave this one too}} \\ &\quad - \frac{h}{2} \underbrace{\frac{d^2}{dh^2}f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)\right)}_{\text{and leave this one too}} \end{aligned}$$

Since we only need $E_2(h)$ and its derivatives at $h = 0$, we don't have to compute the $\frac{d^2f}{dh^2}$ term (thankfully) and we also do not need to compute the $\frac{df}{dh}$ term in E_2' . We do, however, need $\frac{df}{dh}\big|_{h=0}$ for $E_2''(0)$.

$$\begin{aligned} E_2(0) &= \phi(t_n) - y_n = 0 \\ E_2'(0) &= \phi'(t_n) - \frac{1}{2}f(t_n, y_n) - \frac{1}{2}f(t_n, y_n) = 0 \\ E_2''(0) &= \phi''(t_n) - \frac{d}{dh}f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)\right)\bigg|_{h=0} \\ &= \phi''(t_n) - \frac{1}{2}f_t\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)\right)\bigg|_{h=0} \\ &\quad - \frac{1}{2}f(t_n, y_n)f_y\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)\right)\bigg|_{h=0} \\ &= \phi''(t_n) - \frac{1}{2}f_t(t_n, y_n) - \frac{1}{2}f_y(t_n, y_n)f(t_n, y_n) \\ &= \frac{1}{2}\phi''(t_n) \quad \text{by (E9)} \end{aligned}$$

By Taylor, the local truncation error for Euler-2step obeys

Equation D.3.4

$$E_2(h) = \frac{1}{4}\phi''(t_n)h^2 + O(h^3) = \frac{1}{2}Kh^2 + O(h^3) \quad \text{with } K = \frac{1}{2}\phi''(t_n)$$

Observe that the K in (D.3.4) is identical to the K in (D.3.3). This is exactly what we needed in our analysis of Sections D.3.1 and D.3.2.

This page titled [D.3: Variable Step Size Methods](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

E: Hints for Exercises

-

Page ID

91844

This page titled [E: Hints for Exercises](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

F: Answers to Exercises

This page titled [F: Answers to Exercises](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman](#), [Andrew Rechnitzer](#) and [Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

G: Solutions to Exercises

This page titled [G: Solutions to Exercises](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Joel Feldman, Andrew Rechnitzer and Elyse Yeager](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

Index

A

arithmetic mean
[2.2: Averages](#)

C

center of mass
[2.3: Center of Mass and Torque](#)
convergence test
[3.3: Convergence Tests](#)

D

Divergence Test
[3.3: Convergence Tests](#)

F

fundamental theorem of calculus
[1.3: The Fundamental Theorem of Calculus](#)
fundamental theorem of calculus, part 1
[1.3: The Fundamental Theorem of Calculus](#)
fundamental theorem of calculus, part 2
[1.3: The Fundamental Theorem of Calculus](#)

G

geometric mean
[2.2: Averages](#)

H

harmonic mean
[2.2: Averages](#)

I

Integral Test
[3.3: Convergence Tests](#)

M

Midpoint Rule
[1.11: Numerical Integration](#)

N

numerical integration
[1.11: Numerical Integration](#)

R

root mean square
[2.2: Averages](#)

S

sequence
[3.1: Sequences](#)
series
[3.2: Series](#)
Simpson's rule
[1.11: Numerical Integration](#)
squeeze theorem
[3.1: Sequences](#)

T

Taylor remainder
[3.6: Taylor Series](#)
Taylor series
[3.6: Taylor Series](#)
torque
[2.3: Center of Mass and Torque](#)
Trapezoid Rule
[1.11: Numerical Integration](#)
truncation errors
[3.3: Convergence Tests](#)

Glossary

absolute convergence | if the series $\sum_{n=1}^{\infty} |a_n|$ converges, the series $\sum_{n=1}^{\infty} a_n$ is said to converge absolutely

absolute error | if B is an estimate of some quantity having an actual value of A , then the absolute error is given by $|A - B|$

absolute extremum | if f has an absolute maximum or absolute minimum at c , we say f has an absolute extremum at c

absolute maximum | if $f(c) \geq f(x)$ for all x in the domain of f , we say f has an absolute maximum at c

absolute minimum | if $f(c) \leq f(x)$ for all x in the domain of f , we say f has an absolute minimum at c

absolute value function | $f(x) = \begin{cases} -x, & \text{if } x < 0 \\ x, & \text{if } x \geq 0 \end{cases}$

acceleration | is the rate of change of the velocity, that is, the derivative of velocity

acceleration vector | the second derivative of the position vector

algebraic function | a function involving any combination of only the basic operations of addition, subtraction, multiplication, division, powers, and roots applied to an input variable x

alternating series | a series of the form $\sum_{n=1}^{\infty} (-1)^{n+1} b_n$ or $\sum_{n=1}^{\infty} (-1)^n b_n$, where $b_n \geq 0$, is called an alternating series

alternating series test | for an alternating series of either form, if $b_{n+1} \leq b_n$ for all integers $n \geq 1$ and $b_n \rightarrow 0$, then an alternating series converges

amount of change | the amount of a function $f(x)$ over an interval $[x, x+h]$ is $f(x+h) - f(x)$

angular coordinate | θ the angle formed by a line segment connecting the origin to a point in the polar coordinate system with the positive radial (x) axis, measured counterclockwise

antiderivative | a function F such that $F'(x) = f(x)$ for all x in the domain of f is an antiderivative of f

arc length | the arc length of a curve can be thought of as the distance a person would travel along the path of the curve

arc-length function | a function $s(t)$ that describes the arc length of curve C as a function of t

arc-length parameterization | a reparameterization of a vector-valued function in which the parameter is equal to the arc length

arithmetic sequence | a sequence in which the difference between every pair of consecutive terms is the same is called an arithmetic sequence

asymptotically semi-stable solution | $y = k$ if it is neither asymptotically stable nor asymptotically unstable

asymptotically stable solution | $y = k$ if there exists $\epsilon > 0$ such that for any value $c \in (k - \epsilon, k + \epsilon)$ the solution to the initial-value problem $y' = f(x, y), y(x_0) = c$ approaches k as x approaches infinity

asymptotically unstable solution | $y = k$ if there exists $\epsilon > 0$ such that for any value $c \in (k - \epsilon, k + \epsilon)$ the solution to the initial-value problem $y' = f(x, y), y(x_0) = c$ never approaches k as x approaches infinity

autonomous differential equation | an equation in which the right-hand side is a function of y alone

average rate of change | is a function $f(x)$ over an interval $[x, x+h]$ is $\frac{f(x+h) - f(x)}{h}$

average value of a function | (or \bar{f}) the average value of a function on an interval can be found by calculating the definite integral of the function and dividing that value by the length of the interval

average velocity | the change in an object's position divided by the length of a time period; the average velocity of an object over a time interval $[t, a]$ (if $t < a$ or $[a, t]$ if $t > a$), with a position given by $s(t)$, that is $v_{\text{ave}} = \frac{s(t) - s(a)}{t - a}$

base | the number b in the exponential function $f(x) = b^x$ and the logarithmic function $f(x) = \log_b x$

binomial series | the Maclaurin series for $f(x) = (1+x)^r$; it is given by $(1+x)^r = \sum_{n=0}^{\infty} \binom{r}{n} x^n = 1 + rx + \frac{r(r-1)}{2!} x^2 + \dots + \frac{r(r-1)\dots(r-n+1)}{n!} x^n + \dots$ for $|x| < 1$

binormal vector | a unit vector orthogonal to the unit tangent vector and the unit normal vector

boundary conditions | the conditions that give the state of a system at different times, such as the position of a spring-mass system at two different times

boundary point | a point P_0 of R is a boundary point if every δ disk centered around P_0 contains points both inside and outside R

boundary-value problem | a differential equation with associated boundary conditions

bounded above | a sequence $\{a_n\}$ is bounded above if there exists a constant M such that $a_n \leq M$ for all positive integers n

bounded below | a sequence $\{a_n\}$ is bounded below if there exists a constant M such that $M \leq a_n$ for all positive integers n

bounded sequence | a sequence $\{a_n\}$ is bounded if there exists a constant M such that $|a_n| \leq M$ for all positive integers n

cardioid | a plane curve traced by a point on the perimeter of a circle that is rolling around a fixed circle of the same radius; the equation of a cardioid is $r = a(1 + \sin \theta)$ or $r = a(1 + \cos \theta)$

carrying capacity | the maximum population of an organism that the environment can sustain indefinitely

catenary | a curve in the shape of the function $y = a \cosh(x/a)$ is a catenary; a cable of uniform density suspended between two supports assumes the shape of a catenary

center of mass | the point at which the total mass of the system could be concentrated without changing the moment

centroid | the centroid of a region is the geometric center of the region; laminas are often represented by regions in the plane; if the lamina has a constant density, the center of mass of the lamina depends only on the shape of the corresponding planar region; in this case, the center of mass of the lamina corresponds to the centroid of the representative region

chain rule | the chain rule defines the derivative of a composite function as the derivative of the outer function evaluated at the inner function times the derivative of the inner function

change of variables | the substitution of a variable, such as u , for an expression in the integrand

characteristic equation | the equation $a\lambda^2 + b\lambda + c = 0$ for the differential equation $ay'' + by' + cy = 0$

circulation | the tendency of a fluid to move in the direction of curve C . If C is a closed curve, then the circulation of \vec{F} along C is line integral $\int_C \vec{F} \cdot d\vec{s}$, which we also denote $\oint_C \vec{F} \cdot d\vec{s}$.

closed curve | a curve for which there exists a parameterization $\vec{r}(t)$, $a \leq t \leq b$, such that $\vec{r}(a) = \vec{r}(b)$, and the curve is traversed exactly once

closed curve | a curve that begins and ends at the same point

closed set | a set S that contains all its boundary points

comparison test | If $0 \leq a_n \leq b_n$ for all $n \geq N$ and $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges; if $a_n \geq b_n \geq 0$ for all $n \geq N$ and $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.

complementary equation | for the nonhomogeneous linear differential equation $a+2(x)y'' + a_1(x)y' + a_0(x)y = r(x)$, the associated homogeneous equation, called the *complementary equation*, is $a_2(x)y'' + a_1(x)y' + a_0(x)y = 0$

component | a scalar that describes either the vertical or horizontal direction of a vector

component functions | the component functions of the vector-valued function $\vec{r}(t) = f(t)\hat{i} + g(t)\hat{j} + h(t)\hat{k}$ are $f(t)$ and $g(t)$, and the component functions of the vector-valued function $\vec{r}(t) = f(t)\hat{i} + g(t)\hat{j} + h(t)\hat{k}$ are $f(t)$, $g(t)$ and $h(t)$

composite function | given two functions f and g , a new function, denoted $g \circ f$, such that $(g \circ f)(x) = g(f(x))$

computer algebra system (CAS) | technology used to perform many mathematical tasks, including integration

concave down | if f is differentiable over an interval I and f' is decreasing over I , then f is concave down over I

concave up | if f is differentiable over an interval I and f' is increasing over I , then f is concave up over I

concavity | the upward or downward curve of the graph of a function

concavity test | suppose f is twice differentiable over an interval I ; if $f'' > 0$ over I , then f is concave up over I ; if $f'' < 0$ over I , then f is concave down over I

conditional convergence | if the series $\sum_{n=1}^{\infty} a_n$ converges, but the series $\sum_{n=1}^{\infty} |a_n|$ diverges, the series $\sum_{n=1}^{\infty} a_n$ is said to converge conditionally

conic section | a conic section is any curve formed by the intersection of a plane with a cone of two nappes

connected region | a region in which any two points can be connected by a path with a trace contained entirely inside the region

connected set | an open set S that cannot be represented as the union of two or more disjoint, nonempty open subsets

conservative field | a vector field for which there exists a scalar function f such that $\vec{\nabla} f = \vec{F}$

constant multiple law for limits | the limit law $\lim_{x \rightarrow a} c f(x) = c \lim_{x \rightarrow a} f(x) = cL$

constant multiple rule | the derivative of a constant c multiplied by a function f is the same as the constant multiplied by the derivative: $\frac{d}{dx}(cf(x)) = cf'(x)$

constant rule | the derivative of a constant function is zero: $\frac{d}{dx}(c) = 0$, where c is a constant

constraint | an inequality or equation involving one or more variables that is used in an optimization problem; the constraint enforces a limit on the possible solutions for the problem

continuity at a point | A function $f(x)$ is continuous at a point a if and only if the following three conditions are satisfied: (1) $f(a)$ is defined, (2) $\lim_{x \rightarrow a} f(x)$ exists, and (3) $\lim_{x \rightarrow a} f(x) = f(a)$

continuity from the left | A function is continuous from the left at b if $\lim_{x \rightarrow b^-} f(x) = f(b)$

continuity from the right | A function is continuous from the right at a if $\lim_{x \rightarrow a^+} f(x) = f(a)$

continuity over an interval | a function that can be traced with a pencil without lifting the pencil; a function is continuous over an open interval if it is continuous at every point in the interval; a function $f(x)$ is continuous over a closed interval of the form $[a, b]$ if it is continuous at every point in (a, b) , and it is continuous from the right at a and from the left at b

contour map | a plot of the various level curves of a given function $f(x, y)$

convergence of a series | a series converges if the sequence of partial sums for that series converges

convergent sequence | a convergent sequence is a sequence $\{a_n\}$ for which there exists a real number L such that a_n is arbitrarily close to L as long as n is sufficiently large

coordinate plane | a plane containing two of the three coordinate axes in the three-dimensional coordinate system, named by the axes it contains: the xy -plane, xz -plane, or the yz -plane

critical point | if $f'(c) = 0$ or $f'(c)$ is undefined, we say that c is a critical point of f

critical point of a function of two variables | the point (x_0, y_0) is called a critical point of $f(x, y)$ if one of the two following conditions holds: 1. $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$ 2. At least one of $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ do not exist

cross product | $\vec{u} \times \vec{v} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = (u_2v_3 - u_3v_2)\hat{i} - (u_1v_3 - u_3v_1)\hat{j} + (u_1v_2 - u_2v_1)\hat{k}$, where $\vec{u} = (u_1, u_2, u_3)$ and $\vec{v} = (v_1, v_2, v_3)$ determinant a real number associated with a square matrix parallelepiped a three-dimensional prism with six faces that are parallelograms torque the effect of a force that causes an object to rotate triple scalar product the dot product of a vector with the cross product of two other vectors: $\vec{u} \cdot (\vec{v} \times \vec{w})$ vector product the cross product of two vectors.

cross-section | the intersection of a plane and a solid object

cubic function | a polynomial of degree 3; that is, a function of the form $f(x) = ax^3 + bx^2 + cx + d$, where $a \neq 0$

curl | the curl of vector field $\vec{F} = (P, Q, R)$, denoted $\nabla \times \vec{F}$ is the "determinant" of the matrix $\begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}$. ∇ is given by the expression $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$; it measures the tendency of particles at a point to rotate about the axis that points in the direction of the curl at the point

curvature | the derivative of the unit tangent vector with respect to the arc-length parameter

cusp | a pointed end or part where two curves meet

cycloid | the curve traced by a point on the rim of a circular wheel as the wheel rolls along a straight line without slipping

cylinder | a set of lines parallel to a given line passing through a given curve

cylindrical coordinate system | a way to describe a location in space with an ordered triple (r, θ, z) , where (r, θ) represents the polar coordinates of the point's projection in the xy -plane, and z represents the point's projection onto the z -axis

decreasing on the interval I | a function decreasing on the interval I if, for all $x_1, x_2 \in I$; $f(x_1) \geq f(x_2)$ if $x_1 < x_2$

definite integral | a primary operation of calculus; the area between the curve and the x -axis over a given interval is a definite integral

definite integral of a vector-valued function | the vector obtained by calculating the definite integral of each of the component functions of a given vector-valued function, then using the results as the components of the resulting function

degree | for a polynomial function, the value of the largest exponent of any term

density function | a density function describes how mass is distributed throughout an object; it can be a linear density, expressed in terms of mass per unit length; an area density, expressed in terms of mass per unit area; or a volume density, expressed in terms of mass per unit volume; weight-density is also used to describe weight (rather than mass) per unit volume

dependent variable | the output variable for a function

derivative | the slope of the tangent line to a function at a point, calculated by taking the limit of the difference quotient, is the derivative

derivative function | gives the derivative of a function at each point in the domain of the original function for which the derivative is defined

derivative of a vector-valued function | the derivative of a vector-valued function $\vec{r}(t)$ is $\vec{r}'(t) = \lim_{\Delta t \rightarrow 0} \frac{\vec{r}(t + \Delta t) - \vec{r}(t)}{\Delta t}$, provided the limit exists

difference law for limits | the limit law $\lim_{x \rightarrow a} (f(x) - g(x)) = \lim_{x \rightarrow a} f(x) - \lim_{x \rightarrow a} g(x) = L - M$ ∇

difference quotient | of a function $f(x)$ at a is given by $\frac{f(a+h) - f(a)}{h}$ or $\frac{f(x) - f(a)}{x - a}$

difference rule | the derivative of the difference of a function f and a function g is the same as the difference of the derivative of f and the derivative of g : $\frac{d}{dx}(f(x) - g(x)) = f'(x) - g'(x)$

differentiable | a function $f(x, y)$ is differentiable at (x_0, y_0) if $f(x, y)$ can be expressed in the form $f(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) + E(x, y)$, where the error term $E(x, y)$ satisfies $\lim_{(x, y) \rightarrow (x_0, y_0)} \frac{E(x, y)}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} = 0$

differentiable at a | a function for which $f'(a)$ exists is differentiable at a

differentiable function | a function for which $f'(x)$ exists is a differentiable function

differentiable on S | a function for which $f'(x)$ exists for each x in the open set S is differentiable on S

differential | the differential dx is an independent variable that can be assigned any nonzero real number; the differential dy is defined to be $dy = f'(x)dx$

differential calculus | the field of calculus concerned with the study of derivatives and their applications

differential equation | an equation involving a function $y = y(x)$ and one or more of its derivatives

differential form | given a differentiable function $y = f(x)$, the equation $dy = f'(x)dx$ is the differential form of the derivative of y with respect to x

differentiation | the process of taking a derivative

direction angles | the angles formed by a nonzero vector and the coordinate axes

direction cosines | the cosines of the angles formed by a nonzero vector and the coordinate axes

direction field (slope field) | a mathematical object used to graphically represent solutions to a first-order differential equation; at each point in a direction field, a line segment appears whose slope is equal to the slope of a solution to the differential equation passing through that point

direction vector | a vector parallel to a line that is used to describe the direction, or orientation, of the line in space

directional derivative | the derivative of a function in the direction of a given unit vector

directrix | a directrix (plural: directrices) is a line used to construct and define a conic section; a parabola has one directrix; ellipses and hyperbolas have two

discontinuity at a point | A function is discontinuous at a point or has a discontinuity at a point if it is not continuous at the point

discriminant | the value $4AC - B^2$, which is used to identify a conic when the equation contains a term involving xy , is called a discriminant

discriminant | the discriminant of the function $f(x, y)$ is given by the formula $D = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - (f_{xy}(x_0, y_0))^2$

disk method | a special case of the slicing method used with solids of revolution when the slices are disks

divergence | the divergence of a vector field $\vec{F} = (P, Q, R)$, denoted $\nabla \cdot \vec{F}$, is $P_x + Q_y + R_z$; it measures the "outflowing-ness" of a vector field

divergence of a series | a series diverges if the sequence of partial sums for that series diverges

divergence test | if $\lim_{n \rightarrow \infty} a_n \neq 0$, then the series $\sum_{n=1}^{\infty} a_n$ diverges

divergent sequence | a sequence that is not convergent is divergent

domain | the set of inputs for a function

dot product or scalar product | $\vec{u} \cdot \vec{v} = u_1v_1 + u_2v_2 + u_3v_3$ where $\vec{u} = (u_1, u_2, u_3)$ and $\vec{v} = (v_1, v_2, v_3)$

double integral | of the function $f(x,y)$ over the region R in the xy -plane is defined as the limit of a double Riemann sum, $\iint_R f(x,y) \, dA = \lim_{m,n \rightarrow \infty} \sum_{i=1}^m \sum_{j=1}^n f(x_{ij}^*, y_{ij}^*) \, \Delta A$, $\Delta A = \frac{A}{mn}$

double Riemann sum | of the function $f(x,y)$ over a rectangular region $R = \sum_{i=1}^m \sum_{j=1}^n f(x_{ij}^*, y_{ij}^*) \, \Delta A$, where R is divided into smaller subrectangles R_{ij} and (x_{ij}^*, y_{ij}^*) is an arbitrary point in R_{ij}

doubling time | if a quantity grows exponentially, the doubling time is the amount of time it takes the quantity to double, and is given by $(\ln 2)/k$

eccentricity | the eccentricity is defined as the distance from any point on the conic section to its focus divided by the perpendicular distance from that point to the nearest directrix

ellipsoid | a three-dimensional surface described by an equation of the form $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$; all traces of this surface are ellipses

elliptic cone | a three-dimensional surface described by an equation of the form $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$; traces of this surface include ellipses and intersecting lines

elliptic paraboloid | a three-dimensional surface described by an equation of the form $z = \frac{x^2}{a^2} + \frac{y^2}{b^2}$; traces of this surface include ellipses and parabolas

end behavior | the behavior of a function as $x \rightarrow \infty$ and $x \rightarrow -\infty$

epsilon-delta definition of the limit | $\lim_{x \rightarrow a} f(x) = L$ if for every $\epsilon > 0$, there exists a $\delta > 0$ such that if $0 < |x - a| < \delta$, then $|f(x) - L| < \epsilon$

equilibrium solution | any solution to the differential equation of the form $y = c$, where c is a constant

equivalent vectors | vectors that have the same magnitude and the same direction

Euler's Method | a numerical technique used to approximate solutions to an initial-value problem

even function | a function is even if $f(-x) = f(x)$ for all x in the domain of f

explicit formula | a sequence may be defined by an explicit formula such that $a_n = f(n)$

exponent | the value x in the expression b^x

exponential decay | systems that exhibit exponential decay follow a model of the form $y = y_0 e^{-kt}$

exponential growth | systems that exhibit exponential growth follow a model of the form $y = y_0 e^{kt}$

extreme value theorem | if f is a continuous function over a finite, closed interval, then f has an absolute maximum and an absolute minimum

Fermat's theorem | if f has a local extremum at c , then c is a critical point of f

first derivative test | let f be a continuous function over an interval I containing a critical point c such that f is differentiable over I except possibly at c ; if f' changes sign from positive to negative as x increases through c , then f has a local maximum at c ; if f' changes sign from negative to positive as x increases through c , then f has a local minimum at c ; if f' does not change sign as x increases through c , then f does not have a local extremum at c

flux | the rate of a fluid flowing across a curve in a vector field; the flux of vector field \vec{F} across plane curve C is line integral $\int_C \vec{F} \cdot \vec{n} \, ds$

flux integral | another name for a surface integral of a vector field; the preferred term in physics and engineering

focal parameter | the focal parameter is the distance from a focus of a conic section to the nearest directrix

focus | a focus (plural: foci) is a point used to construct and define a conic section; a parabola has one focus; an ellipse and a hyperbola have two

formal definition of an infinite limit | $\lim_{x \rightarrow a} f(x) = \infty$ if for every $M > 0$, there exists a $\delta > 0$ such that if $0 < |x - a| < \delta$, then $f(x) > M$; $\lim_{x \rightarrow a} f(x) = -\infty$ if for every $M > 0$, there exists a $\delta > 0$ such that if $0 < |x - a| < \delta$, then $f(x) < -M$

Frenet frame of reference | (TNB frame) a frame of reference in three-dimensional space formed by the unit tangent vector, the unit normal vector, and the binormal vector

frustum | a portion of a cone; a frustum is constructed by cutting the cone with a plane parallel to the base

Fubini's theorem | if $f(x,y)$ is a function of two variables that is continuous over a rectangular region $R = [a,b] \times [c,d]$, then the double integral of f over the region equals an iterated integral, $\iint_R f(x,y) \, dA = \int_a^b \int_c^d f(x,y) \, dy \, dx = \int_c^d \int_a^b f(x,y) \, dx \, dy$

function | a set of inputs, a set of outputs, and a rule for mapping each input to exactly one output

function of two variables | a function $z = f(x,y)$ that maps each ordered pair (x,y) in a subset D of \mathbb{R}^2 to a unique real number z

Fundamental Theorem for Line Integrals | the value of line integral $\int_C \vec{F} \cdot d\vec{r}$ depends only on the value of f at the endpoints of C : $\int_C \vec{F} \cdot d\vec{r} = f(\vec{r}(b)) - f(\vec{r}(a))$

fundamental theorem of calculus | (also, evaluation theorem) we can evaluate a definite integral by evaluating the antiderivative of the integrand at the endpoints of the interval and subtracting

fundamental theorem of calculus | uses a definite integral to define an antiderivative of a function

fundamental theorem of calculus | the theorem, central to the entire development of calculus, that establishes the relationship between differentiation and integration

general form | an equation of a conic section written as a general second-degree equation

general form of the equation of a plane | an equation in the form $ax + by + cz + d = 0$, where $\vec{n} = (a,b,c)$ is a normal vector of the plane, $P = (x_0, y_0, z_0)$ is a point on the plane, and $d = -ax_0 - by_0 - cz_0$

general solution (or family of solutions) | the entire set of solutions to a given differential equation

generalized chain rule | the chain rule extended to functions of more than one independent variable, in which each independent variable may depend on one or more other variables

geometric sequence | a sequence $\{a_n\}$ in which the ratio a_{n+1}/a_n is the same for all positive integers n is called a geometric sequence

geometric series | a geometric series is a series that can be written in the form $\sum_{n=1}^{\infty} ar^{n-1} = a + ar + ar^2 + ar^3 + \dots$

gradient | the gradient of the function $f(x,y)$ is defined to be $\nabla f(x,y) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$, which can be generalized to a function of any number of independent variables

gradient field | a vector field \vec{F} for which there exists a scalar function f such that $\vec{F} = \nabla f$; in other words, a vector field that is the gradient of a function; such vector fields are also called *conservative*

graph of a function | the set of points (x,y) such that x is in the domain of f and $y = f(x)$

graph of a function of two variables | a set of ordered triples (x,y,z) that satisfies the equation $z = f(x,y)$ plotted in three-dimensional Cartesian space

Green's theorem | relates the integral over a connected region to an integral over the boundary of the region

grid curves | curves on a surface that are parallel to grid lines in a coordinate plane

growth rate | the constant $r > 0$ in the exponential growth function $P(t) = P_0 e^{rt}$

half-life | if a quantity decays exponentially, the half-life is the amount of time it takes the quantity to be reduced by half. It is given by $(\ln 2)/k$

harmonic series | the harmonic series takes the form $\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \dots$

heat flow | a vector field proportional to the negative temperature gradient in an object

helix | a three-dimensional curve in the shape of a spiral

higher-order derivative | a derivative of a derivative, from the second derivative to the n^{th} derivative, is called a higher-order derivative

higher-order partial derivatives | second-order or higher partial derivatives, regardless of whether they are mixed partial derivatives

homogeneous linear equation | a second-order differential equation that can be written in the form $a_2(x)y'' + a_1(x)y' + a_0(x)y = r(x)$, but $r(x) = 0$ for every value of x

Hooke's law | this law states that the force required to compress (or elongate) a spring is proportional to the distance the spring has been compressed (or stretched) from equilibrium; in other words, $F = kx$, where k is a constant

horizontal asymptote | if $\lim_{x \rightarrow \infty} f(x) = L$ or $\lim_{x \rightarrow -\infty} f(x) = L$, then $y = L$ is a horizontal asymptote of f

horizontal line test | a function f is one-to-one if and only if every horizontal line intersects the graph of f , at most, once

hydrostatic pressure | the pressure exerted by water on a submerged object

hyperbolic functions | the functions denoted $\sinh, \cosh, \operatorname{tanh}, \operatorname{csch}, \operatorname{sech}, \operatorname{coth}$, which involve certain combinations of e^x and e^{-x}

hyperboloid of one sheet | a three-dimensional surface described by an equation of the form $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$; traces of this surface include ellipses and hyperbolas

hyperboloid of two sheets | a three-dimensional surface described by an equation of the form $\sqrt{\frac{z^2}{c^2} - \frac{x^2}{a^2} - \frac{y^2}{b^2}} = 1$; traces of this surface include ellipses and hyperbolas

implicit differentiation | is a technique for computing $\frac{dy}{dx}$ for a function defined by an equation, accomplished by differentiating both sides of the equation (remembering to treat the variable y as a function) and solving for $\frac{dy}{dx}$

improper double integral | a double integral over an unbounded region or of an unbounded function

improper integral | an integral over an infinite interval or an integral of a function containing an infinite discontinuity on the interval; an improper integral is defined in terms of a limit. The improper integral converges if this limit is a finite real number; otherwise, the improper integral diverges

increasing on the interval I | a function increasing on the interval I if for all $x_1, x_2 \in I$; $f(x_1) \leq f(x_2)$ if $x_1 < x_2$

indefinite integral | the most general antiderivative of $f(x)$ is the indefinite integral of f ; we use the notation $\int f(x) dx$ to denote the indefinite integral of f

indefinite integral of a vector-valued function | a vector-valued function with a derivative that is equal to a given vector-valued function

independence of path | a vector field \vec{F} has path independence if $\int_{C_1} \vec{F} \cdot d\vec{r} = \int_{C_2} \vec{F} \cdot d\vec{r}$ for any curves C_1 and C_2 in the domain of \vec{F} with the same initial points and terminal points

independent variable | the input variable for a function

indeterminate forms | When evaluating a limit, the forms $\frac{0}{0}, \frac{\infty}{\infty}, 0 \cdot \infty, \infty - \infty, 0^0, \infty^0$, and 1^∞ are considered indeterminate because further analysis is required to determine whether the limit exists and, if so, what its value is.

index variable | the subscript used to define the terms in a sequence is called the index

infinite discontinuity | An infinite discontinuity occurs at a point a if $\lim_{x \rightarrow a^-} f(x) = \pm\infty$ or $\lim_{x \rightarrow a^+} f(x) = \pm\infty$

infinite limit | A function has an infinite limit at a point a if it either increases or decreases without bound as it approaches a

infinite limit at infinity | a function that becomes arbitrarily large as x becomes large

infinite series | an infinite series is an expression of the form $\sum_{n=1}^{\infty} a_n$

inflection point | if f is continuous at c and f changes concavity at c , the point $(c, f(c))$ is an inflection point of f

initial point | the starting point of a vector

initial population | the population at time $t=0$

initial value problem | a problem that requires finding a function y that satisfies the differential equation $\frac{dy}{dx} = f(x)$ together with the initial condition $y(x_0) = y_0$

initial value(s) | a value or set of values that a solution of a differential equation satisfies for a fixed value of the independent variable

initial velocity | the velocity at time $t=0$

initial-value problem | a differential equation together with an initial value or values

instantaneous rate of change | the rate of change of a function at any point along the function a , also called $f'(a)$, or the derivative of the function at a

instantaneous velocity | The instantaneous velocity of an object with a position function that is given by $s(t)$ is the value that the average velocities on intervals of the form $[t, a]$ and $[a, t]$ approach as the values of t move closer to a , provided such a value exists

integrable function | a function is integrable if the limit defining the integral exists; in other words, if the limit of the Riemann sums as n goes to infinity exists

integral calculus | the study of integrals and their applications

integral test | for a series $\sum_{n=1}^{\infty} a_n$ with positive terms a_n , if there exists a continuous, decreasing function f such that $f(n) = a_n$ for all positive integers n , then $\sum_{n=1}^{\infty} a_n$ and $\int_1^{\infty} f(x) dx$ either both converge or both diverge

integrand | the function to the right of the integration symbol; the integrand includes the function being integrated

integrating factor | any function $f(x)$ that is multiplied on both sides of a differential equation to make the side involving the unknown function equal to the derivative of a product of two functions

integration by parts | a technique of integration that allows the exchange of one integral for another using the formula $\int u dv = uv - \int v du$

integration by substitution | a technique for integration that allows integration of functions that are the result of a chain-rule derivative

integration table | a table that lists integration formulas

interior point | a point P_0 of \mathbb{R} is a boundary point if there is a δ disk centered around P_0 contained completely in \mathbb{R}

Intermediate Value Theorem | Let f be continuous over a closed bounded interval $[a, b]$ if z is any real number between $f(a)$ and $f(b)$, then there is a number c in $[a, b]$ satisfying $f(c) = z$

intermediate variable | given a composition of functions (e.g., $f(x(t), y(t))$), the intermediate variables are the variables that are independent in the outer function but dependent on the other variables as well; in the function $f(x(t), y(t))$, the variables x and y are examples of intermediate variables

interval of convergence | the set of real numbers x for which a power series converges

intuitive definition of the limit | If all values of the function $f(x)$ approach the real number L as the values of x approach a , $f(x)$ approaches L

inverse function | for a function f , the inverse function f^{-1} satisfies $f^{-1}(y) = x$ if $f(x) = y$

inverse hyperbolic functions | the inverses of the hyperbolic functions where \cosh and sech are restricted to the domain $[0, \infty)$; each of these functions can be expressed in terms of a composition of the natural logarithm function and an algebraic function

inverse trigonometric functions | the inverses of the trigonometric functions are defined on restricted domains where they are one-to-one functions

iterated integral | for a function $f(x, y)$ over the region R is a $\int_a^b \int_c^d f(x, y) dx dy = \int_a^b \left(\int_c^d f(x, y) dy \right) dx$, $\int_c^d \left(\int_a^b f(x, y) dx \right) dy$, where a, b, c , and d are any real numbers and $R = [a, b] \times [c, d]$

iterative process | process in which a list of numbers $x_0, x_1, x_2, x_3, \dots$ is generated by starting with a number x_0 and defining $x_n = F(x_{n-1})$ for $n \geq 1$

Jacobian | the Jacobian $J(u, v)$ in two variables is a 2×2 determinant $J(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$; the Jacobian $J(u, v, w)$ in three variables is a 3×3 determinant $J(u, v, w) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}$

jump discontinuity | A jump discontinuity occurs at a point a if $\lim_{x \rightarrow a^-} f(x)$ and $\lim_{x \rightarrow a^+} f(x)$ both exist, but $\lim_{x \rightarrow a^-} f(x) \neq \lim_{x \rightarrow a^+} f(x)$

Kepler's laws of planetary motion | three laws governing the motion of planets, asteroids, and comets in orbit around the Sun

Lagrange multiplier | the constant (or constants) used in the method of Lagrange multipliers; in the case of one constant, it is represented by the variable λ

lamina | a thin sheet of material; laminas are thin enough that, for mathematical purposes, they can be treated as if they are two-dimensional

left-endpoint approximation | an approximation of the area under a curve computed by using the left endpoint of each subinterval to calculate the height of the vertical sides of each rectangle

level curve of a function of two variables | the set of points satisfying the equation $f(x, y) = c$ for some real number c in the range of f

level surface of a function of three variables | the set of points satisfying the equation $f(x, y, z) = c$ for some real number c in the range of f

limaçon | the graph of the equation $r = a + b \sin \theta$ or $r = a + b \cos \theta$. If $a = b$ then the graph is a cardioid

limit | the process of letting x or t approach a in an expression; the limit of a function $f(x)$ as x approaches a is the value that $f(x)$ approaches as x approaches a

limit at infinity | a function that approaches a limit value L as x becomes large

limit comparison test | Suppose $a_n, b_n \geq 0$ for all $n \geq 1$. If $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = L \neq 0$, then $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ both converge or both diverge; if $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$ and $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges. If $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \infty$ and $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.

limit laws | the individual properties of limits; for each of the individual laws, let $f(x)$ and $g(x)$ be defined for all $x \neq a$ over some open interval containing a ; assume that L and M are real numbers so that $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} g(x) = M$; let c be a constant

limit of a sequence | the real number L such that a sequence converges is called the limit of the sequence

limit of a vector-valued function | a vector-valued function $\vec{r}(t)$ has a limit \vec{L} as t approaches a if $\lim_{t \rightarrow a} |\vec{r}(t) - \vec{L}| = 0$

limits of integration | these values appear near the top and bottom of the integral sign and define the interval over which the function should be integrated

line integral | the integral of a function along a curve in a plane or in space

linear | description of a first-order differential equation that can be written in the form $a(x)y' + b(x)y = c(x)$

linear approximation | the linear function $L(x) = f(a) + f'(a)(x-a)$ is the linear approximation of f at $x=a$

linear approximation | given a function $f(x,y)$ and a tangent plane to the function at a point (x_0, y_0) , we can approximate $f(x,y)$ for points near (x_0, y_0) using the tangent plane formula

linear function | a function that can be written in the form $f(x) = mx + b$

linearly dependent | a set of functions $f_1(x), f_2(x), \dots, f_n(x)$ for which there are constants c_1, c_2, \dots, c_n , not all zero, such that $c_1 f_1(x) + c_2 f_2(x) + \dots + c_n f_n(x) = 0$ for all x in the interval of interest

linearly independent | a set of functions $f_1(x), f_2(x), \dots, f_n(x)$ for which there are no constants c_1, c_2, \dots, c_n , such that $c_1 f_1(x) + c_2 f_2(x) + \dots + c_n f_n(x) = 0$ for all x in the interval of interest

local extremum | if f has a local maximum or local minimum at c , we say f has a local extremum at c

local maximum | if there exists an interval I such that $f(c) \geq f(x)$ for all $x \in I$, we say f has a local maximum at c

local minimum | if there exists an interval I such that $f(c) \leq f(x)$ for all $x \in I$, we say f has a local minimum at c

logarithmic differentiation | is a technique that allows us to differentiate a function by first taking the natural logarithm of both sides of an equation, applying properties of logarithms to simplify the equation, and differentiating implicitly

logarithmic function | a function of the form $f(x) = \log_b(x)$ for some base $b > 0, b \neq 1$ such that $y = \log_b(x)$ if and only if $b^y = x$

logistic differential equation | a differential equation that incorporates the carrying capacity K and growth rate r into a population model

lower sum | a sum obtained by using the minimum value of $f(x)$ on each subinterval

L'Hôpital's rule | If f and g are differentiable functions over an interval a , except possibly at a , and $\lim_{x \rightarrow a} f(x) = 0 = \lim_{x \rightarrow a} g(x)$ or $\lim_{x \rightarrow a} f(x)$ and $\lim_{x \rightarrow a} g(x)$ are infinite, then $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$, assuming the limit on the right exists or is ∞ or $-\infty$.

Maclaurin polynomial | a Taylor polynomial centered at 0; the n^{th} -degree Taylor polynomial for f at 0 is the n^{th} -degree Maclaurin polynomial for f

Maclaurin series | a Taylor series for a function f at $x=0$ is known as a Maclaurin series for f

magnitude | the length of a vector

major axis | the major axis of a conic section passes through the vertex in the case of a parabola or through the two vertices in the case of an ellipse or hyperbola; it is also an axis of symmetry of the conic; also called the transverse axis

marginal cost | is the derivative of the cost function, or the approximate cost of producing one more item

marginal profit | is the derivative of the profit function, or the approximate profit obtained by producing and selling one more item

marginal revenue | is the derivative of the revenue function, or the approximate revenue obtained by selling one more item

mass flux | the rate of mass flow of a fluid per unit area, measured in mass per unit time per unit area

mathematical model | A method of simulating real-life situations with mathematical equations

mean value theorem | if f is continuous over $[a,b]$ and differentiable over (a,b) , then there exists $c \in (a,b)$ such that $f'(c) = \frac{f(b) - f(a)}{b - a}$

mean value theorem for integrals | guarantees that a point c exists such that $f(c)$ is equal to the average value of the function

method of cylindrical shells | a method of calculating the volume of a solid of revolution by dividing the solid into nested cylindrical shells; this method is different from the methods of disks or washers in that we integrate with respect to the opposite variable

method of Lagrange multipliers | a method of solving an optimization problem subject to one or more constraints

method of undetermined coefficients | a method that involves making a guess about the form of the particular solution, then solving for the coefficients in the guess

method of variation of parameters | a method that involves looking for particular solutions in the form $y_p(x) = u(x)y_1(x) + v(x)y_2(x)$, where y_1 and y_2 are linearly independent solutions to the complementary equations, and then solving a system of equations to find $u(x)$ and $v(x)$

midpoint rule | a rule that uses a Riemann sum of the form $M_n = \sum_{i=1}^n f(m_i) \Delta x$, where m_i is the midpoint of the i^{th} subinterval to approximate $\int_a^b f(x) dx$

minor axis | the minor axis is perpendicular to the major axis and intersects the major axis at the center of the conic, or at the vertex in the case of the parabola; also called the conjugate axis

mixed partial derivatives | second-order or higher partial derivatives, in which at least two of the differentiations are with respect to different variables

moment | if n masses are arranged on a number line, the moment of the system with respect to the origin is given by $M = \sum_{i=1}^n m_i x_i$; if, instead, we consider a region in the plane, bounded above by a function $f(x)$ over an interval $[a,b]$, then the moments of the region with respect to the x - and y -axes are given by $M_x = \rho \int_a^b x f(x) dx$ and $M_y = \rho \int_a^b x f(x) dx$, respectively

monotone sequence | an increasing or decreasing sequence

multivariable calculus | the study of the calculus of functions of two or more variables

nappe | a nappe is one half of a double cone

natural exponential function | the function $f(x) = e^x$

natural logarithm | the function $\ln x = \log_e x$

net change theorem | if we know the rate of change of a quantity, the net change theorem says the future quantity is equal to the initial quantity plus the integral of the rate of change of the quantity

net signed area | the area between a function and the x -axis such that the area below the x -axis is subtracted from the area above the x -axis; the result is the same as the definite integral of the function

Newton's method | method for approximating roots of $f(x)=0$; using an initial guess x_0 ; each subsequent approximation is defined by the equation $x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}$

nonelementary integral | an integral for which the antiderivative of the integrand cannot be expressed as an elementary function

nonhomogeneous linear equation | a second-order differential equation that can be written in the form $a_2(x)y'' + a_1(x)y' + a_0(x)y = r(x)$, but $r(x) \neq 0$ for some value of x

normal component of acceleration | the coefficient of the unit normal vector \vec{N} when the acceleration vector is written as a linear combination of \vec{T} and \vec{N}

normal plane | a plane that is perpendicular to a curve at any point on the curve

normal vector | a vector perpendicular to a plane

normalization | using scalar multiplication to find a unit vector with a given direction

number e | as m gets larger, the quantity $(1 + \frac{1}{m})^m$ gets closer to some real number; we define that real number to be e ; the value of e is approximately 2.71828

numerical integration | the variety of numerical methods used to estimate the value of a definite integral, including the midpoint rule, trapezoidal rule, and Simpson's rule

objective function | the function that is to be maximized or minimized in an optimization problem

oblique asymptote | the line $y = mx + b$ if $f(x)$ approaches it as $x \rightarrow \infty$ or $x \rightarrow -\infty$

octants | the eight regions of space created by the coordinate planes

odd function | a function is odd if $f(-x) = -f(x)$ for all x in the domain of f

one-sided limit | A one-sided limit of a function is a limit taken from either the left or the right

one-to-one function | a function f is one-to-one if $f(x_1) \neq f(x_2)$ if $x_1 \neq x_2$

one-to-one transformation | a transformation $T : G \rightarrow R$ defined as $T(u,v) = (x,y)$ is said to be one-to-one if no two points map to the same image point

open set | a set S that contains none of its boundary points

optimization problem | calculation of a maximum or minimum value of a function of several variables, often using Lagrange multipliers

optimization problems | problems that are solved by finding the maximum or minimum value of a function

order of a differential equation | the highest order of any derivative of the unknown function that appears in the equation

orientation | the direction that a point moves on a graph as the parameter increases

orientation of a curve | the orientation of a curve C is a specified direction of C

orientation of a surface | if a surface has an "inner" side and an "outer" side, then an orientation is a choice of the inner or the outer side; the surface could also have "upward" and "downward" orientations

orthogonal vectors | vectors that form a right angle when placed in standard position

osculating circle | a circle that is tangent to a curve C at a point P and that shares the same curvature

osculating plane | the plane determined by the unit tangent and the unit normal vector

p-series | a series of the form $\sum_{n=1}^{\infty} 1/n^p$

parallelogram method | a method for finding the sum of two vectors; position the vectors so they share the same initial point; the vectors then form two adjacent sides of a parallelogram; the sum of the vectors is the diagonal of that parallelogram

parameter | an independent variable that both x and y depend on in a parametric curve; usually represented by the variable t

parameter domain (parameter space) | the region of the uv -plane over which the parameters u and v vary for parameterization $\langle \mathbf{r}(u,v) = \langle x(u,v), y(u,v), z(u,v) \rangle$

parameterization of a curve | rewriting the equation of a curve defined by a function $y=f(x)$ as parametric equations

parameterized surface (parametric surface) | a surface given by a description of the form $\langle \mathbf{r}(u,v) = \langle x(u,v), y(u,v), z(u,v) \rangle$, where the parameters u and v vary over a parameter domain in the uv -plane

parametric curve | the graph of the parametric equations $x(t)$ and $y(t)$ over an interval $a \leq t \leq b$ combined with the equations

parametric equations | the equations $x=x(t)$ and $y=y(t)$ that define a parametric curve

parametric equations of a line | the set of equations $x=x_0+ta$, $y=y_0+tb$, and $z=z_0+tc$ describing the line with direction vector $\mathbf{v}=\langle a,b,c \rangle$ passing through point (x_0,y_0,z_0)

partial derivative | a derivative of a function of more than one independent variable in which all the variables but one are held constant

partial differential equation | an equation that involves an unknown function of more than one independent variable and one or more of its partial derivatives

partial fraction decomposition | a technique used to break down a rational function into the sum of simple rational functions

partial sum | the k th partial sum of the infinite series $\sum_{n=1}^{\infty} a_n$ is the finite sum $S_k = \sum_{n=1}^k a_n = a_1 + a_2 + a_3 + \dots + a_k$

particular solution | member of a family of solutions to a differential equation that satisfies a particular initial condition

particular solution | a solution $y_p(x)$ of a differential equation that contains no arbitrary constants

partition | a set of points that divides an interval into subintervals

percentage error | the relative error expressed as a percentage

periodic function | a function is periodic if it has a repeating pattern as the values of x move from left to right

phase line | a visual representation of the behavior of solutions to an autonomous differential equation subject to various initial conditions

piecewise smooth curve | an oriented curve that is not smooth, but can be written as the union of finitely many smooth curves

piecewise-defined function | a function that is defined differently on different parts of its domain

planar transformation | a function T that transforms a region G in one plane into a region R in another plane by a change of variables

plane curve | the set of ordered pairs $(f(t),g(t))$ together with their defining parametric equations $x=f(t)$ and $y=g(t)$

point-slope equation | equation of a linear function indicating its slope and a point on the graph of the function

polar axis | the horizontal axis in the polar coordinate system corresponding to $r \geq 0$

polar coordinate system | a system for locating points in the plane. The coordinates are r , the radial coordinate, and θ , the angular coordinate

polar equation | an equation or function relating the radial coordinate to the angular coordinate in the polar coordinate system

polar rectangle | the region enclosed between the circles $r = a$ and $r = b$ and the angles $\theta = \alpha$ and $\theta = \beta$; it is described as $R = \{(r, \theta) \mid a \leq r \leq b, \alpha \leq \theta \leq \beta\}$

pole | the central point of the polar coordinate system, equivalent to the origin of a Cartesian system

polynomial function | a function of the form $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$

population growth rate | is the derivative of the population with respect to time

potential function | a scalar function f such that $\nabla f = \mathbf{F}$

power function | a function of the form $f(x) = x^n$ for any positive integer $n \geq 1$

power law for limits | the limit law $\lim_{x \rightarrow a} (f(x))^n = (\lim_{x \rightarrow a} f(x))^n = L^n$ for every positive integer n

power reduction formula | a rule that allows an integral of a power of a trigonometric function to be exchanged for an integral involving a lower power

power rule | the derivative of a power function is a function in which the power on x becomes the coefficient of the term and the power on x in the derivative decreases by 1: If n is an integer, then $\frac{d}{dx} (x^n) = n x^{n-1}$

power series | a series of the form $\sum_{n=0}^{\infty} c_n x^n$ is a power series centered at $x=0$; a series of the form $\sum_{n=0}^{\infty} c_n (x-a)^n$ is a power series centered at $x=a$

principal unit normal vector | a vector orthogonal to the unit tangent vector, given by the formula $\frac{\mathbf{T}'(t)}{\|\mathbf{T}'(t)\|}$

principal unit tangent vector | a unit vector tangent to a curve C

product law for limits | the limit law $\lim_{x \rightarrow a} (f(x) \cdot g(x)) = \lim_{x \rightarrow a} f(x) \cdot \lim_{x \rightarrow a} g(x) = L \cdot M$

product rule | the derivative of a product of two functions is the derivative of the first function times the second function plus the derivative of the second function times the first function: $\frac{d}{dx} (f(x)g(x)) = f'(x)g(x) + g'(x)f(x)$

projectile motion | motion of an object with an initial velocity but no force acting on it other than gravity

propagated error | the error that results in a calculated quantity $f(x)$ resulting from a measurement error dx

quadratic function | a polynomial of degree 2; that is, a function of the form $f(x) = ax^2 + bx + c$ where $a \neq 0$

quadric surfaces | surfaces in three dimensions having the property that the traces of the surface are conic sections (ellipses, hyperbolas, and parabolas)

quotient law for limits | the limit law $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow a} f(x)}{\lim_{x \rightarrow a} g(x)} = \frac{L}{M}$ for $M \neq 0$

quotient rule | the derivative of the quotient of two functions is the derivative of the first function times the second function minus the derivative of the second function times the first function, all divided by the square of the second function: $\frac{d}{dx} \left(\frac{f(x)}{g(x)} \right) = \frac{f'(x)g(x) - g'(x)f(x)}{(g(x))^2}$

radial coordinate | r the coordinate in the polar coordinate system that measures the distance from a point in the plane to the pole

radial field | a vector field in which all vectors either point directly toward or directly away from the origin; the magnitude of any vector depends only on its distance from the origin

radians | for a circular arc of length s on a circle of radius 1, the radian measure of the associated angle θ is s

radius of convergence | if there exists a real number $R > 0$ such that a power series centered at $x=a$ converges for $|x-a| < R$ and diverges for $|x-a| > R$, then R is the radius of convergence; if the power series only converges at $x=a$, the radius of convergence is $R=0$; if the power series converges for all real numbers x , the radius of convergence is $R=\infty$

radius of curvature | the reciprocal of the curvature

radius of gyration | the distance from an object's center of mass to its axis of rotation

range | the set of outputs for a function

ratio test | for a series $\sum_{n=1}^{\infty} a_n$ with nonzero terms, let $\rho = \lim_{n \rightarrow \infty} |a_{n+1}/a_n|$; if $0 \leq \rho < 1$, the series converges absolutely; if $\rho > 1$, the series diverges; if $\rho = 1$, the test is inconclusive

rational function | a function of the form $f(x) = p(x)/q(x)$, where $p(x)$ and $q(x)$ are polynomials

recurrence relation | a recurrence relation is a relationship in which a term a_n in a sequence is defined in terms of earlier terms in the sequence

region | an open, connected, nonempty subset of \mathbb{R}^2

regular parameterization | parameterization $\langle \mathbf{r}(u,v) = \langle x(u,v), y(u,v), z(u,v) \rangle$ such that $\mathbf{r}_u \times \mathbf{r}_v$ is not zero for point (u,v) in the parameter domain

regular partition | a partition in which the subintervals all have the same width

related rates | are rates of change associated with two or more related quantities that are changing over time

relative error | given an absolute error Δq for a particular quantity, $\frac{\Delta q}{q}$ is the relative error.

relative error | error as a percentage of the actual value, given by $\text{relative error} = \left| \frac{A-B}{A} \right| \cdot 100\%$

remainder estimate | for a series $\sum_{n=1}^{\infty} a_n$ with positive terms a_n and a continuous, decreasing function f such that $f(n) = a_n$ for all positive integers n , the remainder $R_N = \sum_{n=N+1}^{\infty} a_n$ satisfies the following estimate: $\int_{N+1}^{\infty} f(x) dx < R_N < \int_N^{\infty} f(x) dx$

removable discontinuity | A removable discontinuity occurs at a point a if $f(x)$ is discontinuous at a , but $\lim_{x \rightarrow a} f(x)$ exists

reparameterization | an alternative parameterization of a given vector-valued function

restricted domain | a subset of the domain of a function f

riemann sum | an estimate of the area under the curve of the form $\sum_{i=1}^n f(x_i^*) \Delta x$

right-endpoint approximation | the right-endpoint approximation is an approximation of the area of the rectangles under a curve using the right endpoint of each subinterval to construct the vertical sides of each rectangle

right-hand rule | a common way to define the orientation of the three-dimensional coordinate system; when the right hand is curved around the z -axis in such a way that the fingers curl from the positive x -axis to the positive y -axis, the thumb points in the direction of the positive z -axis

RLC series circuit | a complete electrical path consisting of a resistor, an inductor, and a capacitor; a second-order, constant-coefficient differential equation can be used to model the charge on the capacitor in an RLC series circuit

rolle's theorem | if f is continuous over $[a,b]$ and differentiable over (a,b) , and if $f(a) = f(b)$, then there exists $c \in (a,b)$ such that $f'(c) = 0$

root function | a function of the form $f(x) = x^{1/n}$ for any integer $n \geq 2$

root law for limits | the limit law $\lim_{x \rightarrow a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x \rightarrow a} f(x)}$ for all L if n is odd and for $L \geq 0$ if n is even

root test | for a series $\sum_{n=1}^{\infty} a_n$, let $\rho = \lim_{n \rightarrow \infty} \sqrt[n]{|a_n|}$; if $0 \leq \rho < 1$, the series converges absolutely; if $\rho > 1$, the series diverges; if $\rho = 1$, the test is inconclusive

rose | graph of the polar equation $r = a \cos 2\theta$ or $r = a \sin 2\theta$ for a positive constant a

rotational field | a vector field in which the vector at point (x,y) is tangent to a circle with radius $r = \sqrt{x^2 + y^2}$; in a rotational field, all vectors flow either clockwise or counterclockwise, and the magnitude of a vector depends only on its distance from the origin

rulings | parallel lines that make up a cylindrical surface

saddle point | given the function $z = f(x,y)$, the point $(x_0, y_0, f(x_0, y_0))$ is a saddle point if both $f_{xx}(x_0, y_0) = 0$ and $f_{yy}(x_0, y_0) = 0$, but f does not have a local extremum at (x_0, y_0)

scalar | a real number

scalar equation of a plane | the equation $a(x-x_0) + b(y-y_0) + c(z-z_0) = 0$ used to describe a plane containing point $P = (x_0, y_0, z_0)$ with normal vector $n = (a, b, c)$ or its alternate form $ax + by + cz + d = 0$, where $d = -ax_0 - by_0 - cz_0$

scalar line integral | the scalar line integral of a function f along a curve C with respect to arc length is the integral $\int_C f \, ds$, it is the integral of a scalar function f along a curve in a plane or in space; such an integral is defined in terms of a Riemann sum, as is a single-variable integral

scalar multiplication | a vector operation that defines the product of a scalar and a vector

scalar projection | the magnitude of the vector projection of a vector

secant | A secant line to a function $f(x)$ at a is a line through the point $(a, f(a))$ and another point on the function; the slope of the secant line is given by $m_{\text{sec}} = \frac{f(x) - f(a)}{x - a}$

second derivative test | suppose $f'(c) = 0$ and f'' is continuous over an interval containing c ; if $f''(c) > 0$, then f has a local minimum at c ; if $f''(c) < 0$, then f has a local maximum at c ; if $f''(c) = 0$, then the test is inconclusive

separable differential equation | any equation that can be written in the form $y' = f(x)g(y)$

separation of variables | a method used to solve a separable differential equation

sequence | an ordered list of numbers of the form $\{a_1, a_2, a_3, \dots\}$ is a sequence

sigma notation | (also, summation notation) the Greek letter sigma (Σ) indicates addition of the values; the values of the index above and below the sigma indicate where to begin the summation and where to end it

simple curve | a curve that does not cross itself

simple harmonic motion | motion described by the equation $x(t) = c_1 \cos(\omega t) + c_2 \sin(\omega t)$, as exhibited by an undamped spring-mass system in which the mass continues to oscillate indefinitely

simply connected region | a region that is connected and has the property that any closed curve that lies entirely inside the region encompasses points that are entirely inside the region

Simpson's rule | a rule that approximates $\int_a^b f(x) dx$ using the area under a piecewise quadratic function. The approximation S_n to $\int_a^b f(x) dx$ is given by $S_n = \frac{\Delta x}{3} [f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \dots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)]$

skew lines | two lines that are not parallel but do not intersect

slicing method | a method of calculating the volume of a solid that involves cutting the solid into pieces, estimating the volume of each piece, then adding these estimates to arrive at an estimate of the total volume; as the number of slices goes to infinity, this estimate becomes an integral that gives the exact value of the volume

slope | the change in y for each unit change in x

slope-intercept form | equation of a linear function indicating its slope and y -intercept

smooth | curves where the vector-valued function $\vec{r}(t)$ is differentiable with a non-zero derivative

solid of revolution | a solid generated by revolving a region in a plane around a line in that plane

solution curve | a curve graphed in a direction field that corresponds to the solution to the initial-value problem passing through a given point in the direction field

solution to a differential equation | a function $y = f(x)$ that satisfies a given differential equation

space curve | the set of ordered triples $(f(t), g(t), h(t))$ together with their defining parametric equations $x = f(t)$, $y = g(t)$ and $z = h(t)$

space-filling curve | a curve that completely occupies a two-dimensional subset of the real plane

speed | is the absolute value of velocity, that is, $|v(t)|$ is the speed of an object at time t whose velocity is given by $v(t)$

sphere | the set of all points equidistant from a given point known as the *center*

spherical coordinate system | a way to describe a location in space with an ordered triple (ρ, θ, ϕ) , where ρ is the distance between P and the origin ($\rho \geq 0$), θ is the same angle used to describe the location in cylindrical coordinates, and ϕ is the angle formed by the positive z -axis and line segment \overline{OP} , where O is the origin and $0 \leq \phi \leq \pi$

squeeze theorem | states that if $f(x) \leq g(x) \leq h(x)$ for all $x \neq a$ over an open interval containing a and $\lim_{x \rightarrow a} f(x) = L = \lim_{x \rightarrow a} h(x)$ where L is a real number, then $\lim_{x \rightarrow a} g(x) = L$

standard equation of a sphere | $(x-a)^2 + (y-b)^2 + (z-c)^2 = r^2$ describes a sphere with center (a,b,c) and radius r

standard form | the form of a first-order linear differential equation obtained by writing the differential equation in the form $y' + p(x)y = q(x)$

standard form | an equation of a conic section showing its properties, such as location of the vertex or lengths of major and minor axes

standard unit vectors | unit vectors along the coordinate axes: $\hat{i} = (1, 0)$, $\hat{j} = (0, 1)$

standard-position vector | a vector with initial point $(0, 0)$

steady-state solution | a solution to a nonhomogeneous differential equation related to the forcing function; in the long term, the solution approaches the steady-state solution

step size | the increment h that is added to the x value at each step in Euler's Method

Stokes' theorem | relates the flux integral over a surface S to a line integral around the boundary C of the surface S

stream function | if $\vec{F} = (P, Q)$ is a source-free vector field, then stream function g is a function such that $P = g_y$ and $Q = -g_x$

sum law for limits | The limit law $\lim_{x \rightarrow a} (f(x) + g(x)) = \lim_{x \rightarrow a} f(x) + \lim_{x \rightarrow a} g(x) = L + M$

sum rule | the derivative of the sum of a function f and a function g is the same as the sum of the derivative of f and the derivative of g : $\frac{d}{dx} (f(x) + g(x)) = f'(x) + g'(x)$

surface | the graph of a function of two variables, $z = f(x, y)$

surface area | the surface area of a solid is the total area of the outer layer of the object; for objects such as cubes or bricks, the surface area of the object is the sum of the areas of all of its faces

surface area | the area of surface S given by the surface integral $\iint_S dS$

surface independent | flux integrals of curl vector fields are surface independent if their evaluation does not depend on the surface but only on the boundary of the surface

surface integral | an integral of a function over a surface

surface integral of a scalar-valued function | a surface integral in which the integrand is a scalar function

surface integral of a vector field | a surface integral in which the integrand is a vector field

symmetric equations of a line | the equations $\frac{x-x_0}{a} = \frac{y-y_0}{b} = \frac{z-z_0}{c}$ describing the line with direction vector $v = (a, b, c)$ passing through point (x_0, y_0, z_0)

symmetry about the origin | the graph of a function f is symmetric about the origin if $(-x, -y)$ is on the graph of f whenever (x, y) is on the graph

symmetry about the y-axis | the graph of a function f is symmetric about the y-axis if $(-x, y)$ is on the graph of f whenever (x, y) is on the graph

symmetry principle | the symmetry principle states that if a region R is symmetric about a line l , then the centroid of R lies on l

table of values | a table containing a list of inputs and their corresponding outputs

tangent | A tangent line to the graph of a function at a point $(a, f(a))$ is the line that secant lines through $(a, f(a))$ approach as they are taken through points on the function with x -values that approach a ; the slope of the tangent line to a graph at a measures the rate of change of the function at a

tangent line approximation (linearization) | since the linear approximation of f at $x=a$ is defined using the equation of the tangent line, the linear approximation of f at $x=a$ is also known as the tangent line approximation to f at $x=a$

tangent plane | given a function $f(x, y)$ that is differentiable at a point (x_0, y_0) , the equation of the tangent plane to the surface $z=f(x, y)$ is given by $z=f(x_0, y_0)+f_x(x_0, y_0)(x-x_0)+f_y(x_0, y_0)(y-y_0)$

tangent vector | to $\vec{r}(t)$ at $t=t_0$ any vector \vec{v} such that, when the tail of the vector is placed at point $\vec{r}(t_0)$ on the graph, vector \vec{v} is tangent to curve C

tangential component of acceleration | the coefficient of the unit tangent vector \vec{T} when the acceleration vector is written as a linear combination of \vec{T} and \vec{N}

Taylor polynomials | the n^{th} -degree Taylor polynomial for f at $x=a$ is $p_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n$

Taylor series | a power series at a that converges to a function f on some open interval containing a .

Taylor's theorem with remainder | for a function f and the n^{th} -degree Taylor polynomial for f at $x=a$, the remainder $R_n(x) = f(x) - p_n(x)$ satisfies $R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$ for some c between x and a ; if there exists an interval I containing a and a real number M such that $|f^{(n+1)}(x)| \leq M$ for all x in I , then $|R_n(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1}$

telescoping series | a telescoping series is one in which most of the terms cancel in each of the partial sums

term | the number a_n in the sequence $\{a_n\}$ is called the n^{th} term of the sequence

term-by-term differentiation of a power series | a technique for evaluating the derivative of a power series $\sum_{n=0}^{\infty} c_n(x-a)^n$ by evaluating the derivative of each term separately to create the new power series $\sum_{n=1}^{\infty} n c_n(x-a)^{n-1}$

term-by-term integration of a power series | a technique for integrating a power series $\sum_{n=0}^{\infty} c_n(x-a)^n$ by integrating each term separately to create the new power series $C + \sum_{n=0}^{\infty} \frac{c_n}{n+1} (x-a)^{n+1}$

terminal point | the endpoint of a vector

theorem of Pappus for volume | this theorem states that the volume of a solid of revolution formed by revolving a region around an external axis is equal to the area of the region multiplied by the distance traveled by the centroid of the region

three-dimensional rectangular coordinate system | a coordinate system defined by three lines that intersect at right angles; every point in space is described by an ordered triple (x, y, z) that plots its location relative to the defining axes

threshold population | the minimum population that is necessary for a species to survive

total area | total area between a function and the x -axis is calculated by adding the area above the x -axis and the area below the x -axis; the result is the same as the definite integral of the absolute value of the function

total differential | the total differential of the function $f(x, y)$ at (x_0, y_0) is given by the formula $dz = f_x(x_0, y_0)dx + f_y(x_0, y_0)dy$

trace | the intersection of a three-dimensional surface with a coordinate plane

transcendental function | a function that cannot be expressed by a combination of basic arithmetic operations

transformation | a function that transforms a region GG in one plane into a region RR in another plane by a change of variables

transformation of a function | a shift, scaling, or reflection of a function

trapezoidal rule | a rule that approximates $\int_a^b f(x) dx$ using the area of trapezoids. The approximation T_n to $\int_a^b f(x) dx$ is given by $T_n = \frac{\Delta x}{2} [f(x_0) + 2f(x_1) + 2f(x_2) + \dots + 2f(x_{n-1}) + f(x_n)]$.

tree diagram | illustrates and derives formulas for the generalized chain rule, in which each independent variable is accounted for

triangle inequality | If a and b are any real numbers, then $|a+b| \leq |a| + |b|$

triangle inequality | the length of any side of a triangle is less than the sum of the lengths of the other two sides

triangle method | a method for finding the sum of two vectors; position the vectors so the terminal point of one vector is the initial point of the other; these vectors then form two sides of a triangle; the sum of the vectors is the vector that forms the third side; the initial point of the sum is the initial point of the first vector; the terminal point of the sum is the terminal point of the second vector

trigonometric functions | functions of an angle defined as ratios of the lengths of the sides of a right triangle

trigonometric identity | an equation involving trigonometric functions that is true for all angles θ for which the functions in the equation are defined

trigonometric integral | an integral involving powers and products of trigonometric functions

trigonometric substitution | an integration technique that converts an algebraic integral containing expressions of the form $\sqrt{a^2-x^2}$, $\sqrt{x^2+a^2}$, or $\sqrt{x^2-a^2}$ into a trigonometric integral

triple integral | the triple integral of a continuous function $f(x, y, z)$ over a rectangular solid box B is the limit of a Riemann sum for a function of three variables, if this limit exists

triple integral in cylindrical coordinates | the limit of a triple Riemann sum, provided the following limit exists: $\lim_{\Delta r, \Delta \theta, \Delta z \rightarrow 0} \sum_{i=1}^l \sum_{j=1}^m \sum_{k=1}^n f(r_{ijk}^*, \theta_{ijk}^*, s_{ijk}^*) \Delta r_{ijk}^* \Delta \theta_{ijk}^* \Delta z_{ijk}^*$

triple integral in spherical coordinates | the limit of a triple Riemann sum, provided the following limit exists: $\lim_{\Delta \rho, \Delta \varphi, \Delta \theta \rightarrow 0} \sum_{i=1}^l \sum_{j=1}^m \sum_{k=1}^n f(\rho_{ijk}^*, \theta_{ijk}^*, \varphi_{ijk}^*) \Delta \rho_{ijk}^* \Delta \varphi_{ijk}^* \Delta \theta_{ijk}^*$

Type I | a region D in the xy -plane is Type I if it lies between two vertical lines and the graphs of two continuous functions $g_1(x)$ and $g_2(x)$

Type II | a region D in the xy -plane is Type II if it lies between two horizontal lines and the graphs of two continuous functions $h_1(y)$ and $h_2(y)$

unbounded sequence | a sequence that is not bounded is called unbounded

unit vector | a vector with magnitude 1

unit vector field | a vector field in which the magnitude of every vector is 1

upper sum | a sum obtained by using the maximum value of $f(x)$ on each subinterval

variable of integration | indicates which variable you are integrating with respect to; if it is x , then the function in the integrand is followed by dx

vector | a mathematical object that has both magnitude and direction

vector addition | a vector operation that defines the sum of two vectors

vector difference | the vector difference $\vec{v} - \vec{w}$ is defined as $\vec{v} + (-\vec{w}) = \vec{v} + (-1)\vec{w}$

vector equation of a line | the equation $\vec{r} = \vec{r}_0 + t\vec{v}$ used to describe a line with direction vector $\vec{v} = (a, b, c)$ passing through point $P = (x_0, y_0, z_0)$, where $\vec{r}_0 = (x_0, y_0, z_0)$, is the position vector of point P

vector equation of a plane | the equation $\vec{r} \cdot \vec{n} = \vec{r}_0 \cdot \vec{n}$, where P is a given point in the plane, Q is any point in the plane, and \vec{n} is a normal vector of the plane

vector field | measured in \mathbb{R}^2 , an assignment of a vector $\vec{F}(x, y)$ to each point (x, y) of a subset D of \mathbb{R}^2 ; in \mathbb{R}^3 , an assignment of a vector $\vec{F}(x, y, z)$ to each point (x, y, z) of a subset D of \mathbb{R}^3

vector line integral | the vector line integral of vector field \vec{F} along curve C is the integral of the dot product of \vec{F} with unit tangent vector \vec{T} of C with respect to arc length, $\int_C \vec{F} \cdot \vec{T} ds$; such an integral is defined in terms of a Riemann sum, similar to a single-variable integral

vector parameterization | any representation of a plane or space curve using a vector-valued function

vector projection | the component of a vector that follows a given direction

vector sum | the sum of two vectors, \vec{v} and \vec{w} , can be constructed graphically by placing the initial point of \vec{w} at the terminal point of \vec{v} ; then the vector sum $\vec{v} + \vec{w}$ is the vector with an initial point that coincides with the initial point of \vec{v} , and with a terminal point that coincides with the terminal point of \vec{w}

vector-valued function | a function of the form $\vec{r}(t) = f(t)\hat{i} + g(t)\hat{j}$ or $\vec{r}(t) = f(t)\hat{i} + g(t)\hat{j} + h(t)\hat{k}$, where the component functions f , g , and h are real-valued functions of the parameter t .

velocity vector | the derivative of the position vector

vertex | a vertex is an extreme point on a conic section; a parabola has one vertex at its turning point. An ellipse has two vertices, one at each end of the major axis; a hyperbola has two vertices, one at the turning point of each branch

vertical asymptote | A function has a vertical asymptote at $x=a$ if the limit as x approaches a from the right or left is infinite

vertical line test | given the graph of a function, every vertical line intersects the graph, at most, once

vertical trace | the set of ordered triples (c, y, z) that solves the equation $f(c, y) = z$ for a given constant $x=c$ or the set of ordered triples (x, d, z) that solves the equation $f(x, d) = z$ for a given constant $y=d$

washer method | a special case of the slicing method used with solids of revolution when the slices are washers

work | the amount of energy it takes to move an object; in physics, when a force is constant, work is expressed as the product of force and distance

work done by a force | work is generally thought of as the amount of energy it takes to move an object; if we represent an applied force by a vector \vec{F} and the displacement of an object by a vector \vec{s} , then the work done by the force is the dot product of \vec{F} and \vec{s} .

zero vector | the vector with both initial point and terminal point $(0,0)$

zeros of a function | when a real number x is a zero of a function f , $f(x) = 0$

δ ball | all points in \mathbb{R}^3 lying at a distance of less than δ from (x_0, y_0, z_0)

δ disk | an open disk of radius δ centered at point (a, b)

Detailed Licensing

Overview

Title: [CLP-2 Integral Calculus \(Feldman, Rechnitzer, and Yeager\)](#)

Webpages: 75

Applicable Restrictions: Noncommercial

All licenses found:

- [CC BY-NC-SA 4.0](#): 88% (66 pages)
- [Undeclared](#): 12% (9 pages)

By Page

- [CLP-2 Integral Calculus \(Feldman, Rechnitzer, and Yeager\)](#) - [CC BY-NC-SA 4.0](#)
 - [Front Matter](#) - [Undeclared](#)
 - [TitlePage](#) - [Undeclared](#)
 - [InfoPage](#) - [Undeclared](#)
 - [Table of Contents](#) - [Undeclared](#)
 - [Colophon](#) - [CC BY-NC-SA 4.0](#)
 - [Licensing](#) - [Undeclared](#)
 - [Preface](#) - [CC BY-NC-SA 4.0](#)
 - [Dedication](#) - [CC BY-NC-SA 4.0](#)
 - [Acknowledgements](#) - [CC BY-NC-SA 4.0](#)
 - [Using the exercises in this book](#) - [CC BY-NC-SA 4.0](#)
 - [Feedback about the text](#) - [CC BY-NC-SA 4.0](#)
 - [1: Integration](#) - [CC BY-NC-SA 4.0](#)
 - [1.1: Definition of the Integral](#) - [CC BY-NC-SA 4.0](#)
 - [1.2: Basic properties of the definite integral](#) - [CC BY-NC-SA 4.0](#)
 - [1.3: The Fundamental Theorem of Calculus](#) - [CC BY-NC-SA 4.0](#)
 - [1.4: Substitution](#) - [CC BY-NC-SA 4.0](#)
 - [1.5: Area between curves](#) - [CC BY-NC-SA 4.0](#)
 - [1.6: Volumes](#) - [CC BY-NC-SA 4.0](#)
 - [1.7: Integration by parts](#) - [CC BY-NC-SA 4.0](#)
 - [1.8: Trigonometric Integrals](#) - [CC BY-NC-SA 4.0](#)
 - [1.9: Trigonometric Substitution](#) - [CC BY-NC-SA 4.0](#)
 - [1.10: Partial Fractions](#) - [CC BY-NC-SA 4.0](#)
 - [1.11: Numerical Integration](#) - [CC BY-NC-SA 4.0](#)
 - [1.12: Improper Integrals](#) - [CC BY-NC-SA 4.0](#)
 - [1.13: More Integration Examples](#) - [CC BY-NC-SA 4.0](#)
 - [2: Applications of Integration](#) - [CC BY-NC-SA 4.0](#)
 - [2.1: Work](#) - [CC BY-NC-SA 4.0](#)
 - [2.2: Averages](#) - [CC BY-NC-SA 4.0](#)
 - [2.3: Center of Mass and Torque](#) - [CC BY-NC-SA 4.0](#)
 - [2.4: Separable Differential Equations](#) - [CC BY-NC-SA 4.0](#)
 - [3: Sequence and series](#) - [CC BY-NC-SA 4.0](#)
 - [3.1: Sequences](#) - [CC BY-NC-SA 4.0](#)
 - [3.2: Series](#) - [CC BY-NC-SA 4.0](#)
 - [3.3: Convergence Tests](#) - [CC BY-NC-SA 4.0](#)
 - [3.4: Absolute and Conditional Convergence](#) - [CC BY-NC-SA 4.0](#)
 - [3.5: Power Series](#) - [CC BY-NC-SA 4.0](#)
 - [3.6: Taylor Series](#) - [CC BY-NC-SA 4.0](#)
 - [3.7: Optional — Rational and irrational numbers](#) - [CC BY-NC-SA 4.0](#)
 - [Appendices](#) - [CC BY-NC-SA 4.0](#)
 - [A: High School Material](#) - [CC BY-NC-SA 4.0](#)
 - [A.1: Similar Triangles](#) - [CC BY-NC-SA 4.0](#)
 - [A.2: Pythagoras](#) - [CC BY-NC-SA 4.0](#)
 - [A.3: Trigonometry — Definitions](#) - [CC BY-NC-SA 4.0](#)
 - [A.4: Radians, Arcs and Sectors](#) - [CC BY-NC-SA 4.0](#)
 - [A.5: Trigonometry — Graphs](#) - [CC BY-NC-SA 4.0](#)
 - [A.6: Trigonometry — Special Triangles](#) - [CC BY-NC-SA 4.0](#)
 - [A.7: Trigonometry — Simple Identities](#) - [CC BY-NC-SA 4.0](#)
 - [A.8: Trigonometry — Add and Subtract Angles](#) - [CC BY-NC-SA 4.0](#)
 - [A.9: Inverse Trigonometric Functions](#) - [CC BY-NC-SA 4.0](#)
 - [A.10: Areas](#) - [CC BY-NC-SA 4.0](#)
 - [A.11: Volumes](#) - [CC BY-NC-SA 4.0](#)
 - [A.12: Powers](#) - [CC BY-NC-SA 4.0](#)
 - [A.13: Logarithms](#) - [CC BY-NC-SA 4.0](#)
 - [A.14: Highschool Material You Should be Able to Derive](#) - [CC BY-NC-SA 4.0](#)
 - [A.15: Cartesian Coordinates](#) - [CC BY-NC-SA 4.0](#)
 - [A.16: Roots of Polynomials](#) - [CC BY-NC-SA 4.0](#)
 - [B: Complex Numbers and Exponentials](#) - [CC BY-NC-SA 4.0](#)
 - [B.1: Definition and Basic Operations](#) - [CC BY-NC-SA 4.0](#)

- B.2: The Complex Exponential - *CC BY-NC-SA 4.0*
- C: More About Numerical Integration - *CC BY-NC-SA 4.0*
 - C.1: Richardson Extrapolation - *CC BY-NC-SA 4.0*
 - C.2: Romberg Integration - *CC BY-NC-SA 4.0*
 - C.3: Adaptive Quadrature - *CC BY-NC-SA 4.0*
- D: Numerical Solution of ODE's - *CC BY-NC-SA 4.0*
 - D.1: Simple ODE Solvers — Derivation - *CC BY-NC-SA 4.0*
 - D.2: Simple ODE Solvers — Error Behaviour - *CC BY-NC-SA 4.0*
 - D.3: Variable Step Size Methods - *CC BY-NC-SA 4.0*
- Back Matter - *Undeclared*
 - E: Hints for Exercises - *CC BY-NC-SA 4.0*
 - F: Answers to Exercises - *CC BY-NC-SA 4.0*
 - G: Solutions to Exercises - *CC BY-NC-SA 4.0*
 - Index - *Undeclared*
 - Glossary - *Undeclared*
 - Detailed Licensing - *Undeclared*